ABSTRACT

Title of Dissertation: PROMOTING RICH AND LOW-BURDEN SELF-TRACKING WITH MULTIMODAL DATA INPUT

Yuhan Luo
Doctor of Philosophy, 2022

Dissertation Directed by: Associate Professor Eun Kyoung Choe
College of Information Studies

Manual tracking of personal data offers many benefits such as increased engagement and situated awareness. However, existing self-tracking tools often employ touch-based input to support manual tracking, imposing a heavy input burden and limiting the richness of the collected data. Inspired by speech's fast and flexible nature, this dissertation examines how speech input works with traditional touch input to manually capture personal data in different contexts: food practice, productivity, and exercise.

As a first step, I conducted co-design workshops with registered dietitians to explore opportunities for customizing food trackers composed of multimodal input. The workshops generated diverse tracker designs to meet dietitians' information needs, with a wide range of tracking items, timing, data format, and input modalities.

In the second study, I specifically examined how speech input supports capturing everyday food practice. I created FoodScrap, a speech-based food journaling app, and conducted a data collection study, in which FoodScrap not only collected rich details of meals and food decisions, but was also recognized for encouraging self-reflection.

To further integrate touch and speech on mobile phones, I developed NoteWordy, a multimodal system integrating touch and speech input to capture multiple types of data. Through deploying NoteWordy in the context of productivity tracking, I found several input patterns varying by the data type as well as participants' input habits, error tolerance, and social surroundings. Additionally, speech input helped faster entry completion and enhanced the richness of the free-form text.

Furthermore, I expanded the research scope by exploring speech input on smart speakers by developing TandemTrack, a multimodal exercise assistant coupling a mobile app and an Alexa skill. In a four-week deployment study, TandemTrack demonstrated the convenience of the hands-free speech input to capture exercise data and acknowledged the importance of visual feedback on the mobile app to help with data exploration.

Across these studies, I describe the strengths and limitations of speech as an input modality to capture personal data in various contexts, and discuss opportunities for improving the data capture experience with natural language input. Lastly, I conclude the dissertation with design recommendations toward a low-burden, rich, and reflective self-tracking experience.

PROMOTING RICH AND LOW-BURDEN SELF-TRACKING WITH MULTIMODAL DATA INPUT

by

Yuhan Luo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Associate Professor Eun Kyoung Choe, Chair/Advisor
Dr. Bongshin Lee (Microsoft Research)
Assistant Professor Hernisa Kacorri
Associate Professor Beth St. Jean
Professor Philip Resnik, Dean’s Representative

Dedication

To my mom, for her never-ending love and support.

Acknowledgments

It is said that when we have completed a journey and all else has passed, that is the time to get nostalgic about it. As my five-year Ph.D. study comes to its end, I can finally revel in the good old days and express my deepest gratitude to those who have supported me along the way.

First and foremost, I feel extremely fortunate to have worked with my advisor, Eun Kyoung Choe, who opened up a whole new world of HCI research for me seven years ago and bravely took me on as her first Ph.D. student. Over the years, Eun Kyoung has been unwearyingly mentoring and supporting me with hands-on exercises on research, teaching, and writing. Her rigorous research spirit, critical mindset, and meticulous work ethic influenced me profoundly, which apply more than academic work. More importantly, she always encouraged me to go beyond what I think I am capable of, from which I have learned to challenge myself and be open to new possibilities. Even when I was upset and lost, she placed trust and confidence in me, and never hesitated to offer emotional support. I would not be where I am today, without her encouragement and inspiration.

I appreciate all the thoughtful guidance and feedback from my dissertation committee members: Bongshin Lee, Hernisa Kacorri, Beth St Jean, and Philip Resnik. Thanks to Bongshin, who has been mentoring me on several exciting research projects since I was a master student. Her candid and sharp opinions often presented me with thought-provoking questions, which motivated me to dive deep and keep exploring new directions. Thanks to Hernisa, who always shared valuable advice on research and career. With her expertise in accessibility and inclusive design, Hernisa's insights into supporting the disability community with multimodal interaction will no doubt help shape my future work. Thanks to Beth, who has been incredibly patient and always has her door open for questions. She constantly ensured that my research addressed potential concerns of health consumers and reviewed all my writings in detail. Thanks to Philip, who generously served as the dean's representative as well as a voting member of my dissertation committee. His feedback on my literature review and data analysis were extremely helpful, which nudged me to articulate my thoughts in a more logical and powerful way.

The completion of my research would not have been possible without many collaborators. I would like to thank Young-Ho Kim, who built OmniTrack Research, an amazing platform that enabled two of the studies in my dissertation. His highly organized design and programming habits also inspired me to improve my own work. In the meantime, I would like to thank Peiyi Liu, for helping with the co-design study; Daniel Smolyak, for developing the mobile app version of TandemTrack; Naeemul Hassan, for helping write the FoodScrap paper; Haritha Malladi and Shwetank Jain, for helping with the qualitative coding in the NoteWordy project; and Chi Young Oh, for collaborating on a clinical survey and interview study outside my dissertation work. I also want to thank all my study participants for their interests and contributions. A special thanks to Jared Strange, the best writing fellow ever, who carefully proofread many of my papers and dissertation chapters.

At UMD, it has been a pleasure to study and work with a group of like-minded friends: Annie Li, my statistics pro and workout buddy, who was open to discuss anything with me and always reminded me that I am more than my work; Xiaoyun Huang and Xin Qian, with whom I explored many interesting places in the DMV area; and Di Hu, who prepared and shared so many great foods with me. I am also grateful to my peers at HCIL and THE Lab, including Alisha Pradhan, Emma Dixon, Diana Chou, Jarrett Lee, Jong Ho Lee, Lily Huang, Kyungyeon Lee, Pramod Chundury, and Yuting Liao, for exchanging their thoughts and fun facts in life. In addition, I owe thanks to professors Amanda Lazar, Catherine Plaisant, Mega Subramaniam, Niklas Elmqvist, and Jesscia Vitak, for their valuable feedback on my research work, papers, and presentations.

I am deeply grateful for my experience at Penn State, where I made the decision to pursue a Ph.D. in HCI, even though I did not know what it meant just a few months before. I would like to thank professors Chao-Hsien Chu, David Conroy, Luke Zhang, and Mary Beth Rosson, who helped with my master's thesis and Ph.D. application; and the smart and hardworking friends, Hanqing Zhao, Jun Ge, Na Sun, Yan Wang, Qi Jin, and Can Zhang, with whom I spent the great two years.

I was fortunate to develop friendships with many wonderful people, even though we are in different places around the world. I would like to thank Xinning Gui and Yubo Kou, for actively sharing their experiences as senior academic fellows and for having fun with me as close friends. Outside academia, I am grateful to many life-long friends with whom I have grown for more than ten years, especially Shu Hao, Xi Zhu, Jie Zhao, and Wenlu Zhang, who always make themselves available to me whenever I want to talk, confide, or travel around. Additional thanks go to Jiahuan Liu, Meng Gao, Mengwei Zhang, Ren Nan, Yuqing Shi, Yuanjing Chen, Yingying Zhou, and Zhe Cha. Our joyful dancing, happy laughter, and beautiful trips through nature will never be forgotten, and I look forward to our reunion in the near future.

Last, but not least, I am forever grateful to my family, especially my mom, Jiaojiao Bi, for making education a priority and for unconditionally supporting everything I do. As I moved further and further away from home, her love and encouragement never diminished. I would also like to thank my dad, Yundong Luo, for getting me interested in computing technology as a kid and for always backing me behind; my uncle, Yafei Bi, and aunt, Hongbo Lu, for being the family role models since my childhood and for warmly taking care of me during my stay in the United States. A hearty thanks to my chubby cat, Missy, for bringing endless comfort and joy to me.

Table of Contents

List of Tables

List of Figures

Glossary

Term	Definition
Adherence	The extent to which an individual performas the target behavior in accordance with a consistent frequency.
Data capture burden	The burden that capturing personal data imposes to people, including time, physical/cognitive load, emotional burden, learning effort, and privacy concerns, etc.
Data field	An entry where people can enter data. Depending on the data processing need, a data field can be designed in different formats, such as number, Likert scale, date and time, and text.
Data richness	The level of details included in the unstructured data (e.g., text) to explain specific questions or themes.
Design widget	A miniature design component that can be used to compose an larger interface or system. Design widgets can be paper-based or digitized, depending on the goals of the design activity.
In-situ data capture	Data capture situated in the original time and place.
Multimodal input	User input through two or more combined modes, such as speech, touch, gesture, and motion. This dissertation mainly focuses on speech and touch input.
Quantified-Selfer	A community that share personal stories of self-tracking.
Regimen	A prescribed plan such as capturing certain types of data (data capture regimen) and performing a target activity (e.g., exercise regimen).

Term	Definition
Self-reflection	A way of carefully thinking about or assessing one's own behavior or beliefs.
Self-tracking	A practice of recording the occurrences of one's target behavior or thoughts, which is also described as self-monitoring.
Self-tracker	An individual who practices self-tracking.
Streak	The number of days of doing an activity in a row.
Tracker	A system, app, or device that allows people to capture specific data about themselves, such as food tracker, mood tracker, and sleep tracker.
Utterance	An uninterrupted chain of spoken or written language. In this dissertation, an utterance refers to spoken language.

Chapter 1: Introduction

1.1 Motivation

Self-tracking, a practice of noticing and recording the occurrence of one's target behavior [1,2], has become increasingly popular among individuals. Believing in the notion of "*gaining self-knowledge through numbers*," people practice self-tracking to improve their health or other aspects of life and to explore new experience that satisfies their curiosity [3,4]. As of 2019, 42% of Americans reported tracking one or more health metrics about themselves with digital technologies [5]. While lots of data can be captured automatically with embedded sensors and wearable devices [6], manual tracking of personal data still affords many benefits, such as increased engagement [7] and situated awareness [1]. In the context of food tracking, for example, the action of capturing food intake provides people an opportune moment to reflect on their eating behaviors [8]. This reflective process can further encourage positive eating intentions resulting in a healthier diet [8]. In addition, manual tracking allows people to collect personal contexts (e.g., social environments) that are difficult to automatically capture, which are important for healthcare providers to develop personalized treatment for their patients [9, 10].

However, existing tools often employ screen-based touch input to facilitate manual tracking [11, 12], which can be limited in collecting rich contextual information (e.g., selecting an item from multiple choices) or impose a heavy data capture burden (e.g., typing). On the other hand, the recent development of speech interaction, with its fast and flexible nature [13, 14], has created opportunities to support more efficient manual tracking. First, speech input can lower the input burden of unstructured text data because people normally speak faster than they type [15]. Second, with spoken language, people can describe their activities and thoughts in a flexible manner (e.g., different ways to execute time queries [16]), which can reduce the mental load required to organize or rephrase specific information. Third, people tend to talk expressively [14, 17], and therefore generate rich details that might otherwise be overlooked with touch input. Lastly, interaction with some voice-activated devices is hands-free (e.g., Amazon Echo [18]), which further lowers the data capture burden by allowing people to capture their data without the need to look at the device screen.

Despite these positive characteristics, speech input is not perfect. For example, speech input is limited in supporting people to edit their data on the fly [19]; it can also raise privacy concerns in public settings [20]. In these cases, touch input can be helpful for data editing and mitigating privacy concerns. Given that self-tracking covers a wide range of activities and multiple data types [21], current research provides little understanding of how speech and touch input can complement each other to optimize people's data capture experience. For example, when each input modality is preferred for which types of data, how they can be combined in different tracking contexts, and whether the incorporation of speech input can help lower the data capture burden and enhance data richness.

Therefore, my dissertation examines how speech input can work with the traditional touch input to manually collect personal data, focusing on people's ***modality preferences***,

data capture burden and ***data richness***. Taking a mixed-methods approach, I situate people to use speech input in several self-tracking contexts, including food practice, productivity, and exercise, which are three essential aspects of human health and wellbeing [22,23]. My overarching goal is to provide a deep understanding of the strengths and limitations of speech input in collecting different types of structured and unstructured data, so that we can best leverage speech input's potential and avoid design pitfalls in building effective multimodal self-tracking systems.

In this dissertation, the term ***structured data*** refers to data that have predefined format and length with certain patterns [24]. Examples of structured data include dates, timestamps, numbers, and domain-specific items (e.g., food name with its quantity). ***Unstructured data***, on the other hand, is "*everything else*" that does not follow a specific format [24]. Examples of unstructured data are text, audio, and video.

1.2 Thesis Statement

My thesis claims are summarized in the following statements:

To maximize the benefits of manual tracking, technology should afford easy data capture and promote data richness. Speech as a fast and natural input modality can complement the limitations of traditional touch input by lowering the input burden whilst collecting rich details in unstructured data. On the other hand, speech input's main drawbacks, including lack of editing support and situational constraints, can be compensated by touch input. Therefore, a multimodal self-tracking system that effectively combines touch and speech input can help individuals take advantage of *both modalities, so as to easily collect rich data.*

1.3 Research Questions and Approaches

To verify the thesis statement, my dissertation consists of four strands of empirical research studies aiming to answer five research questions. In the following, I categorize the research questions into three themes:

Identifying design opportunities for multimodal self-tracking tools

RQ1. What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems?

Integrating speech and touch input on mobile phones to support self-tracking

RQ2. What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?

RQ3. How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes?

RQ4. How does the input modality affect the data richness in unstructured data?

Examining the values of a smart speaker in supporting consistent self-tracking

RQ5. How does a smart speaker complement and augment a mobile app in supporting consistent exercise?

Research Question	Addressed in	Domain	Data Types	Target audience	Input modalities
RQ1. What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems?	**Chapter 4.** Co-designing with registered dietitians	Food practice	*Structured* (e.g., calorie) and *unstructured* data (e.g., reflection thoughts)	Healthcare providers	*Any input* on paper-based prototypes
RQ2. What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?	**Chapter 5.** FoodScrap: a data collection study	Food practice	*Unstructured* data (e.g., food decisions)	Self-trackers	*Speech input* on mobile phone
RQ3. How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes? **RQ4.** How does the input modality affect the data richness in unstructured input?	**Chapter 6.** NoteWordy: a design and deployment study	Productivity	*Structured* (e.g., task duration) and *unstructured* data (e.g., feelings)	Self-trackers	*Touch* and *speech input* on mobile phone
RQ5. How does a smart speaker complement and augment a mobile app in supporting consistent exercise?	**Chapter 7.** TandemTrack: a design, deployment, and comparison study	Exercise	*Structured* data (workout repetitions)	Self-trackers	*Touch input* on mobile phone, *speech input* on smart speaker

Table 1.1: A summary of the research questions and how they were addressed, along with corresponding tracking domain, target audience, data types, and input modalities.

To answer these five research questions, I conducted a co-design study involving healthcare providers to explore design opportunities for multimodal data input to customize food trackers (RQ1). Next, I specifically examined how speech input supports people to capture their daily food practices (RQ2). With the lessons learned, I designed and developed a multimodal self-tracking system integrating both touch and speech input to capture different types of data (RQ3 and RQ4). Furthermore, I expanded the scope of speech input from mobile phones to smart speakers, examining how people use different input modalities across devices (RQ5). Table 1.1 summarizes the research questions and how they are addressed.

In this dissertation, I use the term *system*, *tool*, and *application* interchangeably to refer to any technology which people engage with. I refer the action of systematically collecting one's personal data to *tracking*, *journaling*, *data capture*, or *data collection*. A *tracker* refers to a system or device that allows people to capture specific data about themselves, such as food tracker, mood tracker, and sleep tracker. For an individual who practices self-tracking but does not belong to a specific population (e.g., patients), I denote them as a *self-tracker*.

1.4 Contributions

This dissertation makes the following methodological, empirical, and artifact contributions to the Human-Computer Interaction (HCI) and Personal Informatics fields:

1. Structuring co-design workshops involving healthcare providers to solicit design considerations of self-tracking tools for patients with different dietary problems. Unlike prior co-design studies that often focused on designing one specific interface, our co-design study generated diverse tracker designs resulting from patients' characteristics and dietitians' practice styles. By detailing the procedure of the co-design workshops, I illustrate how to contextualize both researchers and providers in patients' experience and foster providers' creativity.

2. Identification of multiple dimensions from healthcare providers' perspective to customize food trackers for different patients, including tracking items, data format, timing and frequency of tracking. These customization dimensions informed design opportunities for creating tracking templates to collect clinical relevant data and promote patient-provider collaboration.

3. Design, development, and evaluation of NoteWordy, a multimodal mobile app in-

tegrating touch and speech input to capture both structured and unstructured data in different types. The data collection with NoteWordy showed how people used the two input modalities in different scenarios and demonstrated how speech input can lower the data capture burden while promoting data richness of free-form text.

4. Design, development, and evaluation of a multimodal system TandemTrack that couples a mobile app and an Alexa skill to support exercise training and tracking. The deployment study with TandemTrack detailed how people used touch and speech input across devices for different exercise features and informed design opportunities for supporting consistent exercise through developing an integrated multimodal exercise experience.

5. Empirical understandings based on field deployment, questionnaires, and interviews, to reveal the strengths and limitations of speech input in three important self-tracking contexts: food practice, productivity, and exercise. Such understandings shed light on design considerations for combining the two input modalities to optimize people's data capture experience with multimodal data input.

1.5 Thesis Overview

This dissertation centers around how multimodal data input supports capturing different types of data for self-tracking purposes, and is divided into eight chapters:

Chapter 2, Background, reviews prior research in self-tracking and highlight the benefits that manual tracking affords. I also describe prior works that incorporated speech input to collect personal data and summarize the research gaps in existing literature.

Chapter 3, Research Contexts and Methodological Foundation, in which I describe the three self-tracking contexts that I examine in this dissertation—food practice, productivity, and exercise—with related academic research and commercial applications. In addition, I provide the methodology foundations of how self-tracking systems are designed and evaluated in HCI research.

Chapter 4, Design Opportunities for Multimodal Input to Customize Food Trackers, in which I report findings from six individual co-design workshops with registered dietitians, detailing dietitians' information needs and diverse food tracker design for patients with various dietary problems.

Chapter 5, Understanding How Speech Input Supports Food Journaling, presents a one-week data collection study examining how people capture their everyday food practice, including food components, preparation methods, and food decisions via speech input. The study showed the promise of speech input in collecting rich eating contexts and enabling situated reflection, and informed opportunities for better leveraging the large amount of speech input data to support data exploration.

Chapter 6, Design and evaluation of NoteWordy, presents how I designed, developed, and evaluated a multimodal self-tracking application NoteWordy that integrates touch and speech input. Through deploying NoteWordy in the context of productivity tracking for two weeks, I found how people used the two input modalities to capture different types of data, and how their modality choices were affected by their previous input habit, error tolerance, and social surroundings. I also show how speech input reduced time spent on completing the diary entries and enhanced data richness in free-form text.

Chapter 7, Designing a Multimodal Exercise Tracking System, presents a comparison study with TandemTrack, a multimodal exercise assistant coupling a mobile app and an Alexa skill (a speech-activated application on Amazon Echo devices). During the study, one group of participants used both the app and the skill to perform a four-week exercise challenge and other group used the mobile app only. I describe the factors influencing people's modality choices for performing exercise and reviewing feedback and illustrate how multimodal interaction enriched their exercise experience.

Chapter 8, Summary and Future Work, reflects on the findings across the four studies and summarizes the research approaches and contributions, as well as future research opportunities.

Chapter 2: Background

My dissertation builds upon prior research on self-tracking, particularly the importance of manual tracking to promote engagement in data capture and enhance situated awareness. In this chapter, I first provide an overview of self-tracking, including its purposes and applications in our daily life and different approaches to practice self-tracking. Next, to explore design opportunities for making manual tracking easier, I examine the potential of speech input by reviewing previous work on speech interaction and Natural Language Interfaces (NLIs) that were used for data capture purposes. Lastly, I summarize the key takeaways from the literature and describe how they inspire my dissertation.

2.1 Self-Tracking

Self-tracking, also known as self-monitoring, refers to "*an individual noticing and recording the occurrences of his or her own target behaviors*" [1]. Self-tracking helps people collect a wide range of data, from objective facts (e.g., step counts, heart rate, blood pressure) to subjective assessments (e.g., symptoms, mood, inner thoughts) [3, 7]. Over the past few years, we have witnessed a growing interest among individuals, such as the Quantified-Self (QS) community, in tracking and exploring data about themselves [3]. As of 2019, 42% of Americans reported tracking one or more health metrics using digital technologies [5]. In 2021, over 400,000 mobile apps were found to support personal data tracking in app stores [25]. While the majority of these data focus on physical health, some pertain to other aspects of life such as work productivity and emotional wellbeing [26, 27]. No matter what data are tracked, the ultimate goal of self-tracking often centers on improving life quality by gaining deeper self-understanding.

2.1.1 Purposes and Applications of Self-Tracking

There has been theoretical research on self-tracking since the 1970s, mostly in cognitive and behavioral science [1, 28]. Two main purposes of self-tracking—***assessment*** and ***therapeutic***—and their applications in our life are described below.

2.1.1.1 Assessment Purposes: Sharing Personal Health Data at Clinics

Traditionally, self-tracking was used for ***assessment*** purposes in clinical settings [2], where the data being tracked allows clinicians to assess patients' conditions and determine treatment strategies. For example, to treat patients with dietary problems (e.g., obesity, diabetics), clinicians often need to refer to patients' food intake, including calories and nutrients [29, 30]; to develop more personalized care, some clinicians also collect other types of data from patients, such as their water intake and physical activities [31]. Similarly, Cognitive Behavioral Therapy for Insomnia (CBI-i) has been found useful to help patients manage sleep disorders by identifying and treating the factors influencing patients' sleep behavior [32]. To employ CBT-i, clinicians ask patients to track their sleep habits and daytime activities, so that they can provide tailored therapy to individual patients [9]. In the past, these personal data were collected through questionnaires [33], interviews [34], or paper-based diaries [35, 36]. Today, many people turn to digital tracking tools such as mobile apps [37] or wearable devices [38] for convenience.

With the increasing availability of tracking devices and applications, hospitals and other stakeholders in the healthcare industry (e.g., health insurance companies) have been investing in technologies to incorporate personal health and behavioral data into their systems [39, 40]. However, sharing personal health data at clinics is not always helpful. Clinicians may not be able to make use of the data due to lack of time [9]; they may also feel overwhelmed by the large amount of data that are not standardized or do not meet their information needs [41]. Taking patients with irritable bowel syndrome (IBS) as an example, clinicians often need to assess patients' IBS symptoms and food intake with precise timestamps [41, 42]. However, existing food tracking apps primarily focus on calorie counting, failing to provide the information that clinicians need [41]. Thus, to support effective data sharing between clinicians and patients, it is important to take clinicians' work flow and their information needs into account.

2.1.1.2 Therapeutic purposes: Encouraging Positive Behaviors

In addition to assessment purposes, research has shown that consistently tracking a target behavior can lead to ***reactive effect***, which will change the frequency of the behavior [1]. The reactive effect can be explained through increased self-awareness on the target behavior, which makes individuals think about the behavior more frequently and seriously; and eventually change the behavior if it departs from their personally acceptable standard or the social norm [43]. Oftentimes, people gain awareness of their behavior either through the feedback loop of self-tracking or consistently engaging in the data collection process [43, 44]. As such, researchers have incorporated different strategies to design self-tracking applications as behavior change interventions [26, 45, 46]. In 2006, before the mobile phone era, Lin and colleagues created Fish'n'Step, a social computer game that promoted physical activities by linking people's step counts to the growth of an animated virtual character [46]. Through a 14-week long deployment study, researchers found that the social game served as a catalyst for promoting exercise and improving players' attitudes towards physical activity [46]. More recently, Kim and colleagues developed TimeAware, a desktop widget to boost work productivity by tracking and displaying information workers' productivity data with different visual designs [26]. Researchers found that emphasizing non-productive activities significantly improved workers' productivity level, but such effect was not sustainable: workers' productivity significantly decreased later when the intervention was withdrawn [26]. In addition, self-tracking has been applied to discourage negative behaviors. For example, Paay and Colleagues built QuittyLink, a mobile app that tracks people's smoking activities and provides a set of visualizations on their smoking data, which helped participants form strategies to resist the smoking desire and reduce smoking behavior [47].

However, behavior change is a long and complex process, with a high chance of relapse [48]. To evaluate if a system can led to behavior change, researchers often need to conduct large-scale and long-term studies; even so, it is still difficult to tell whether the observed change is sustainable [49]. Prior research and challenges in evaluating behavior change systems is described in Section 3.4.2.

2.1.2 Self-Tracking Approaches

People use various self-tracking technologies to capture their personal data automatically (e.g., accelerometers [6, 50]), manually (e.g., food diary apps [51, 52]), or a combination of both (e.g., automatically capture sleep from wearable devices while manually entering sleep quality [21]). In the following, I describe these approaches and the input modalities that are involved.

2.1.2.1 Automated Tracking

With the exponential growth of wearable devices (e.g., Fitbit [50]), embedded sensors in home objects (e.g., Smart Pillbox [53]), and accelerometers built in mobile phones (e.g., Google Fit [54]), we have witness an increasing popularity of automated tracking technologies. Automated tracking lowers the data input burden and increases data accuracy [6], but not all the data can be captured automatically, especially our subjective assessment such as feelings and thoughts. Some the wearable devices can be cumbersome to wear [55], and home sensing technologies may impose privacy concerns [56]. In addition, automated tracking reduces individuals' involvement in data collection process, making it difficult for them to understand and make sense of their data, especially when there is a lack of other contextual information (e.g., automatically detected stress level [57]). As a result, self-awareness, let alone self-reflection for potential behavior change, can be hardly improved [57].

Main Input Modality	Research / Commercial Apps	Data Types
Screen-based touch / click	SleepTight [a] [11]	*Structured*: bed time (time), sleep quality (multiple choice), coffee intake (number and time), etc
	MyFitnessPal [b] [52]	*Structured*: food name & quantities (domain-specific items), nutrients & calorie consumption (number)
	Break Tracking [a] [58]	*Structured*: break-taking time (time), productivity & mood (Likert scale)
Photo [c]	OneNote Meal [a] [8]	*Unstructured*: food (raw photo), food healthiness assessment (free-form text)
	Bitesnap [b] [59]	*Structured*: food name (domain-specific items extracted from photos), calorie consumption (number)
Speech	Journify [b] [60]	*Unstructured*: daily thoughts (raw audios)
	Talk-to-Track [b] [61]	*Structured*: food name & exercise type (domain specific items extracted from speech input), calorie consumption / burn (number)
Handwriting	Bullet Journal [a] [7, 62]	*Unstructured*: anything (free-form writing & drawing)

Table 2.1: Representative HCI and Health Informatics research / commercial applications for manual tracking with different input modalities.

[a]Research prototype
[b]Commercial application
[c]Different from passive motion sensing, the action of photo tracking is initiated by self-trackers.

2.1.2.2 Manual Tracking

Unlike automated tracking that predominately focuses on capturing structured and objective data with predefined format, manual tracking offers people the flexibility to capture unstructured and subjective data [3, 7, 10, 63, 64], such as self-reported health symptoms, mood, and inner thoughts. More importantly, people play an active role in making sense of their data in the process of manual tracking, leading to enhanced self-awareness and reactive effect [65]. Such effects could further enable situated reflection and increase the intention of behavior change [2, 8, 37]. For example, my colleagues and I conducted a diary study on photo-based meal tracking in 2018 [8], in which two groups of participants recorded the meals that they considered healthy or unhealthy with text explanation. We found that the action of capturing one's meals and assessing the meal healthiness made participants more cautious about their food choices: those who tracked healthy meals were encouraged to maintain a healthy diet, and those who tracked unhealthy meals were deterred from eating unhealthy food [8]. Table 2.1 summarizes representative research work as well as commercial applications for manual tracking using different approaches.

However, the heavy data capture burden of manual tracking makes long-term manual tracking challenging [3, 12, 66–68]. In the context of tracking physical activities, for example, manually logging all the activity sessions is burdensome, which can lead to missing entries or inaccurate information [69, 70]. Similarly, in practicing food journaling, it is difficult to deliberately log all the food details such as meal composition, condiment, and preparation methods [12]. To date, limited solutions have been developed to reduce the data capture burden of manual tracking, among which the focus is still on screen-based touch input [11, 71]. Although a few applications have incorporated other input modalities such as speech and photo-taking, they often focus on capturing only a single type of data (e.g., Bitesnap [59], Journify [60]). More details about speech-based self-tracking applications are described in Section 2.2.

2.1.2.3 Semi-Automated Tracking and Beyond

In 2014, Choe's dissertation "*Designing Self-Monitoring Technology to Promote Data Capture and Reflection*" sought to support easy and accurate manual tracking while promoting self-reflection to nudge positive behavior changes [72]. Her work leveraged the mobile phone's lock screen to lower the data entry barrier and applied framing effects to create persuasive feedback. Through understanding the benefits and limitations of manual tracking using touch input, Choe proposed "semi-automated tracking," a balanced approach combining automated and manual tracking to optimize data capture burden and reflection support [73]. Later in 2019, Kim's dissertation "*Designing Flexible Self-Tracking Technologies for Enhancing In Situ Data Collection Capability*" expanded Choe's work by developing flexible tools that are generative and adaptive to cover diverse self-tracking contexts [74]. In particular, Kim employed the semi-automated tracking approach to build OmniTrack [21], which allows people to customize what data to track and which format to use. With OmniTrack, people can create their own trackers by configuring different data fields (e.g., time, number, Likert scale, multiple choice, free-form text), and then manually enter the data; they can also import data from other automated tracking devices or applications (e.g., Fitbit [50], Google Fit [54]).

Although semi-automated tracking leveraged the strengths of both automated and manual tracking to lower the data capture burden while engaging people in data collection, the manual tracking part still largely focused on screen-based touch input [21, 75]. My dissertation therefore explores opportunities for using speech as an input modality to support manual tracking. The main questions include whether and how speech input can make manual tracking easier or enhance the data richness, and how people choose between or combine speech and touch input in various self-tracking contexts. With these questions, the following reviews existing research on speech-based data collection.

2.2 Speech-Based Data Collection

Speech is one the most commonly used communication channels for people to interact with one and another [76]. With the advance in speech recognition [77] and natural language processing (NLP) [78] technology, we see numerous applications incorporating speech as an input modality, ranging from speech-to-text services (e.g., live transcription [79], voice typing [80]) to natural language interfaces (NLI) that respond to user intent (e.g., speech-enabled data analytics [16, 81]). The recent introduction of intelligent voice assistants embedded in mobile phones and smart speakers (e.g., Amazon Alexa [82], Google Assistant [83], and Apple Siri [84]) further drives the adoption of speech interaction for everyday use [85]. In this section, I first cover prior research that employed speech input for personal data collection, and then describe existing work on natural language interfaces (NLIs), focusing on how they are applied in self-tracking contexts.

2.2.1 Personal Data Capture With Speech Input

Compared with manual typing, speech input is faster and more flexible [15], allowing people to quickly interact with a digital system using expressions they are familiar with (e.g., time-related queries [16]). Therefore, researchers have begun to build data collection tools such as survey instruments with speech input to increase the cost-effectiveness [17, 86–88]. For example, Revilla and colleagues compared speech and text input in responding to survey questions, and found that participants who used speech input spent less time completing their responses and provided more elaborate answers (i.e., answers with additional descriptive information or explanation about a theme) than those who typed [17]. In Patnaik and colleagues' study on collecting self-reported health symptoms, the researchers found when patients used speech recordings, they provided more accurate information than when they type in SMS messages or fill electronic forms [89]. Although not all about speech input, Zhang and colleagues designed Eat4Thoughts, which supports people to capture their eating activities through video recording [90]. Researchers found that the audio elements in the videos complemented the visual images in providing rich contextual information about one's eating experience, and the free-form expression also encouraged people to reflect on eating behaviors in-situ [90].

However, speech input is not always ideal for capturing personal data. One limitation of speech input is the difficulty with editing when mistakes were made [19]. In addition, there is a general impression that speech interface is vulnerable to recognition errors—once an error occurs, it can be difficult to recover [91]. As a result, people who have negative experiences using speech input may resist using it again [91]. Another limitation is related to capturing personal data in public spaces: some people feel embarrassed talking to their phone in front of others or concerned about their private information being disclosed [20]. Therefore, instead of putting speech and touch input in competition, my work integrates touch and speech as a whole so that people can take advantage of both input modalities to capture their data in the way they like.

2.2.2 Natural Language Interfaces (NLIs) for Data Capture

The rapid development of natural language processing (NLP) technologies has accelerated the rise of Natural Language Interfaces (NLIs), in which words, phrases, or sentences act as commands for creating, selecting, and modifying data [92–94]. To interpret and execute these commands, NLIs employ both rule-based approaches and machine learning techniques, which aid in processing the unstructured language sources into structured objects such as entity, time, and event [95–97]. A typical example is the reminder service on voice assistants (e.g., Amazon Alexa [82], Google Assistant [83]): people can set up a reminder by speaking to the device and saying "*remind me at 8 am every day to exercise*," from which the service automatically extracts the time, event, and whether this is a recurring reminder. Due to the complicated sentiment and ambiguity in natural languages, current NLIs are not yet generalizable to handle all kinds of input [78]. Even the state-of-the-art systems need either a knowledge base (e.g., vocabularies, syntax rules) or domain-specific training data to effectively perform language processing [96, 97].

Among the self-tracking applications with speech input, many only keep the original audio recordings or transcribed text (e.g., Journify [60], Day One [98], Murmur [99]), which makes it difficult for people to retrieve the key information. Although NLIs were found to be easy and convenient in supporting various activities (e.g., ask a question from Google Assistant [83]), it has not drawn the attention of Personal Informatics researchers until recently. In 2018, a commercial app Talk-to-Track was launched [61], allowing people to capture their food intake and exercise with spoken language, from which it automatically extracts the food or exercise type and calculates the calorie consumption or burn [61]. However, to accurately differentiate multiple food items or exercise activities, Talk-to-Track requires people to deliberately separate each item by saying "*comma*" [61], which limits the flexibility of speech input. In 2019, Korpusik and colleagues built Coco Nutrition, a conversational calorie counter that detects more fine-grained food information (e.g., portion size, nutrients) from both written text and speech input [100]. In a similar vein, Silva and colleagues implemented ModEat in 2021, a multimodal food journaling app that supports food recognition from text, database search, barcode, and speech input [101]. Leveraging an external NLP service, ModEat can recognize a variety of foods and calculate people's calorie consumption [101].

While demonstrating the promises of NLIs to support personal data capture, existing work predominantly focused on capturing only a single type of data such as numbers or domain-specific items (with food name and quantity being the most common). In real-world settings, however, people often capture multiple types of data (e.g., time, location, event, and free-form notes) about their daily activities [21], which is not well-supported in current speech-based data capture systems. To bridge the gap between NLIs and speech-based self-tracking, I aim to leverage speech input's fast and flexible data capture as well as helping people easily review the key information in their data.

2.3 Chapter 2 Summary

In this Chapter, I covered the theoretical background of self-tracking and prior research on speech-based data collection. Learning about the variety of input modalities, I was motivated to start my dissertation by ***identifying design opportunities for multimodal self-tracking tools*** (Research Theme 1). Inspired by the assessment function of self-tracking, my goal is to collect valid and useful data that meet healthcare providers' information needs.

The benefits and limitations of manual tracking motivated me to design more efficient manual tracking tools. Through reviewing existing applications and research work on speech-based data collection, I see the underexplored potential of speech to facilitate data capture and the promises of natural language processing (NLP) to better organize different types of data. Therefore, I aim to ***integrate speech and touch input on mobile phones to support self-tracking*** (Research Theme 2).

Furthermore, the prevalence of speech interaction encouraged me to expand my research scope from mobile phones to other devices; the therapeutic function of self-tracking also encouraged me to explore the opportunities for speech input to facilitate positive behavior change. In this light, I aim to ***examine the values of a smart speaker in supporting consistent self-tracking*** (Research Theme 3).

Chapter 3: Research Contexts and Methodological Foundations

To address the three research themes [1] of my dissertation, I situate people to use multimodal input in three different self-tracking contexts—food practice, productivity, and exercise—because (1) they are essential aspects of human health and wellbeing [22, 23]; and (2) these contexts involve tracking different types of personal data (both structured and unstructured) that fit my research goals. In this chapter, I first review prior research and commercial applications for tracking food practice, productivity, and exercise data. I then describe methodology foundations in HCI for designing and evaluating self-tracking technology, and how they inspire my research design.

3.1 Food Journaling

3.1.1 Food Journaling For Assessment Purposes

Food journaling originated in clinical settings, where healthcare providers assess patients' food consumption to understand their nutrient intake and eating habits [35]. Traditional food assessment methods include interviews (e.g., 24-hour food recall [34]), questionnaires (e.g., food frequency questionnaire (FFQ) [33]), and paper-based diaries (e.g.,

[1] Theme 1: Identifying design opportunities for multimodal self-tracking tools
Theme 2: Integrating speech and touch input on mobile phones to support self-tracking
Theme 3: Examining the values of a smart speaker in supporting consistent self-tracking

dietary record form [36]). Different food assessment methods serve for different purposes. For example, 24-hour food recall assesses people' most recent food intake and dietary history focuses on understanding their long-term eating patterns [35]. As people's food practices vary by their age, culture, and personal preferences, it is common for providers to modify existing food assessment methods in order to suit people with different diets. For example, there are various versions of food frequency questionnaires (FFQ), such as FFQ for children [102] and FFQ for women [103].

With the advances in mHealth technology, many commercial applications (e.g., MyFitnessPal [52], Lose It [104], MyFoodDiary [51], Bitesnap [59], YouAte [105]) are marketed to support food journaling. Although these applications outperform traditional food assessment methods with better adherence [106] and higher accuracy [107], most of them focus on weight loss and calorie watching, which are not appropriate for those with specific tracking needs [41,42,108–110]. As an example, the most important goal for irritable bowel syndrome (IBS) patients is to identify the food triggers that cause their IBS symptoms [41, 108]. However, even with the tools that support self-experimentation through tracking food and symptoms, the level of details needed for each patient differ from one another because of their different lifestyle and stage of the syndrome [41, 108]. Furthermore, Eikey and colleagues found that women suffering from eating disorders are subject to obsessive logging because many food journaling tools afford detailed calorie counting [109, 110]. Such design could prompt patients to obsess over the number of their calorie intake, to the point that their eating disorders are exacerbated. The premise of my first study (Chapter 4) is the mismatch between existing food tracker design and the diverse tracking needs of individuals and their healthcare providers.

3.1.2 Collecting Rich Eating Contexts

While prior works primarily focused on capturing food components with the aim of providing more accurate nutrients and calorie information, we see a growing interest in Human-Food Interaction (HFI), which focuses on enriching food practice ranging from how people cook, how they interact with food, and how food influences their daily life [111, 112]. In particular, the practice of food journaling has been expanded to capture broader eating contexts (e.g., mood, eating environment) beyond just what foods people eat [8, 90, 113–115]. Research also highlighted the importance of enabling situated reflection on one's eating behavior [8, 90, 116], which is seen as one of the leading reasons for behavior change [117]. For example, Zhang and colleagues developed Eat4Thought, which captures a variety of eating contexts such as mood, emotion, and eating environment via video recording and manually tagging the food characteristics [90]. Through vivid documentation and reflection on one's own eating experience, Eat4Thought raised people's healthy eating awareness and prompted them to think about how external factors such as social relationships influenced their eating behaviors [90]. To further leverage the rich data of people's eating contexts, Terzimehić and colleagues collected a set of food choice moments, including food photos, preparation methods, level of hunger, eating occasion, and food choice rationale [114]. These data characterized a wide range of real-life eating events, which informed opportunities for designing personalized healthy eating interventions [114]. For the ease of data analysis, prior works often predefined certain structures in terms of what data to capture and which format to use (e.g., selecting a mood from existing options), but little work has looked into what people capture about their food choices in unstructured forms, how rich the information is, and how much data capture burden the action imposes on people.

3.1.3 Challenges of Food Journaling

Despite the benefits in supporting clinical assessment and enabling meaningful self-reflection, food journaling is also known to be burdensome due to the complexity of meal composition and variation in preparation methods [12]. We have seen research effort made to lower the burden of food journaling using different input modalities, including photo-based food journals [118], barcode scanning [119], accelerated search [120], and smart sensors [121], among which photo-based food journal is most popular for its convenience [118] and ease of sharing [113]. Because food photos may include information such as location and social elements, they further reduce the burden of capturing additional eating contexts [107]. However, food photos cannot always capture necessary details such as portion size, condiments, and individual ingredients, especially when the meal preparation methods are not visually obvious [8, 10]. In addressing such challenges, Chung and colleagues designed Foodprint, a mobile app that allows people to add contextual information (e.g., mood, symptoms, free-form text descriptions) in addition to food photos to aid situated reflection [10]. In addition, researchers have built systems to automatically collect food information, such as a body-worn wearable tracker with motion sensors [122] and an earphone-like device utilizing microphone signals [121]. Although their approaches reduced input burden and improved data accuracy, it may undermine the benefits of in-the-moment awareness created by food journaling.

3.2 Productivity Tracking

3.2.1 Enhancing Productivity Through Self-Tracking

According to World Health Organization (WHO), productivity is one of the most important factors that greatly influence the quality of life [123]. Regardless of what kind of job people do, they are always looking for ways to boost productivity, and feel depressed when their productivity decreases [123]. In the field of Personal Informatics, researchers have designed and built numerous self-tracking applications to help people enhance their productivity [23,26,124–129]. Some of these applications focused on restricting the use of digital devices (e.g., Forest [128], MyTime [125]), and others focused on helping people understand their time spent (e.g., TimeAware [26], RescueTime [129], TimeCamps [130]). For example, Forest helps people reduce mobile phone usage by "growing" different aesthetic virtual trees as reward for non-screen time [128], and RescueTime automatically calculates a "productivity score" according to one's time spent on different applications and websites [129]. White and colleagues proposed a framework that guides people to design a full cycle of self-tracking activities for better productivity [124]. In the framework, they highlighted how organizational factors (e.g., work environment, job characteristics) and individual factors (e.g., motivation, lifestyle) play a part in productivity, and demonstrated how collecting and analyzing these data can enable reflection and behavior change [124].

3.2.2 Collecting Contextualized Personal Productivity Data

While productivity is often measured by time spent or work output, Kim and colleagues have identified multiple dimensions that people use to conceptualize their productivity, including task achievement, enjoyment, long-term career benefit, social and spiritual benefit, and emotional status [131]. As such, the concept "producvity" may have different meanings for different people. To better understand how people define and evaluate their productivity, a commonly used research approach is to collect productivity data situated in real-life settings as ecologically valid sources [23, 131, 132]. For example, Mark and colleagues collected information workers' mood and productivity level on a daily basis in their offices and identified several factors contributing to workplace well-being, including sleep quality and email usage [23]. More recently, Cao and colleagues examined individuals' multi-tasking behavior in remote meetings through a large-scale diary study, through which they revealed how often multi-tasking occurred and identified both positive and negative consequences on productivity resulting from different ways of multi-tasking [132].

In addition to work-related activities, researchers also highlighted the roles of breaks, which can significantly affect work productivity [133]. Hence, researchers have looked into the contexts around how people take breaks during work hours [58, 134–136]. For example, targeting information workers, Epstein and colleagues categorized different types of breaks (e.g., social break, physical break, outdoor break) through a diary study, and showed how these breaks influenced workers' productivity [58]. To examine the gap between information workers' break-taking intentions and practices, my colleagues and

I collected their planned breaks and actual breaks for three weeks, during which participants acknowledged the importance of breaks but also highlighted the conflicts between taking breaks and maintaining productivity [134]. The insights learned from these studies can help researchers and organizations better manage workplace ergonomics and support a healthier work-life balance.

3.3 Exercise Tracking

3.3.1 Mobile Fitness Apps

Recent years have witnessed an increasing interest in developing mobile applications to assist daily exercise. As of March 2020, over 71,000 fitness apps have been launched globally in both the Apple store and Google Play Store [137]. Although fitness tracking apps are shown effective in encouraging physical activity, many people still find it challenging to exercise consistently due to lack of time and environmental constraints [138–142], and fail to meet the recommended level of exercise [143].

To lower exercise barriers, researchers and designers have incorporated behavior change strategies into fitness apps, including exercise guidance [144, 145], data capture [145], performance feedback [144–146], and reminders [147]. Popular fitness training apps such as Nike Training Club [148], SWORKIT [149], 30 Day Fitness Challenge [150], and Keep Trainer [151], provide hundreds of exercise guidance to people with different fitness goals, ranging from yoga, stretch, to high-intensity training. The guidance often takes people through a series of exercise sessions, accompanied by video instructions to demonstrate proper postures. To ensure that people follow the guidance precisely, researchers also explored the feasibility of using 3D techniques to illustrate and correct people's body movements in detail [152, 153]. In addition, many fitness apps capture people's exercise data and provide feedback on their progress and performance. For example, the home screen of 30 Day Fitness Challenge [150] shows a graph of one's weight change and exercise progress, aiming to motivate those who want to lose weight to exercise more. Nike Training Club [148] summarizes individuals' total workouts, highlights their achieved milestones, and provides a list of workout history. To aid people in achieving their exercise goals, researchers have designed stylized representations (e.g., virtual characters) of one's physical activities to indicate their exercise performance [45, 46]. To help people build a regular exercise routine, apps such as Keep Trainer [151], allow people to set their personalized exercise schedule, and send them daily reminders.

Although not part of health and fitness apps, conversational agents (CAs) have been used to promote self-awareness and reflect on one's health and activity data [154, 155]. Commercial apps such as Lark [156] and HealthyBot [157] serve as a "Chatbot," actively initiating conversations with people by asking about their daily activities and well-being. By processing the natural language input and generating responses as feedback, the CAs present opportunities to engage people with their personal health data in new ways [154]. However, such communication requires people to manually input their activity information, which can add extra interaction burden [154].

Despite being equipped with behavior change strategies, most fitness apps provide only single interaction modality, that is, screen-based touch input on a mobile phone. My dissertation, on the other hand, aims to explore how multiple input modalities on other devices can complement mobile apps in supporting exercise tracking, as a way to expand existing mHealth infrastructure [158].

3.3.2 Speech-Based Exercise Assistants

With the introduction of smart speakers such as Amazon Echo [18] and Google Home [159], the market for speech-based applications (e.g., Alexa skills [160], Google Actions [161]) has been expanding rapidly. The latest statistics showed that more than 100,000 Alexa skills and 33,000 Google actions—the equivalent of mobile apps on smartphones—are available in the United States [162]. In 2017, researchers found that only 309 Alexa skills and Google actions were listed under the "Health & Fitness" category [163]; but in 2021, the number of "health & fitness" Alexa skills alone has reached over 2,200 [164]. Among these applications, many support fitness training by providing audio-based exercise regimens (e.g., 7-Minute Workout [165], 5-Minute Plank [166]). For instance, 7-Minute Workout [165] guides people through 14 sets of workout sessions, during which they can pause the session by saying "*pause*" or start the session by saying "*Ready.*" With Amazon Echo devices, the hands-free interaction makes it easy to follow the exercise regimen, because people can focus on their body postures without visual distraction. However, most of these speech-based applications lack support for data capture and exercise feedback—a valuable means to engage people with their exercise data. A few health & fitness mobile apps have a skill version (e.g., Fitbit skill [167], MyFitnessPal skill [168]), allowing people to ask questions about their data captured by the companion apps (e.g., "*How much calories did I burn yesterday?*" "*Did I log my weight today?*"), the skill version of these apps do not support data capture using speech.

In addition, prior research has explored opportunities for multimodal interfaces to support health and fitness activities utilizing other physical devices [169, 170]. Turunen and colleagues created a health and fitness companion consisting of a mobile app and an intelligent agent that was designed like a little rabbit [169]. While their work aimed at optimizing the accuracy of speech recognition by providing daily health advice [169], I examine how people use the mobile app and the speech input on the smart speaker for in-home exercise training and tracking.

3.4 Technology Design and Evaluation in HCI Research

This dissertation follows the spirit of human-centered design, which includes empathizing with target users, defining key questions, iterating on design ideas, prototyping, and evaluation [171, 172]. In this section, I focus on explaining the research methods I chose to design and evaluate the proposed systems (Chapter 4–7). I first describe how HCI researchers collect design requirements through co-designing with stakeholders, and then illustrate how we examine a self-tracking system's potential benefits and assess feasibility of real-world adoption.

3.4.1 Collecting Design Requirements Through Co-Designing With Stakeholders

Co-design, also called participatory design, is a collaborative activity in which stakeholders are empowered to actively participate in the stage of product design and development [173–176]. In HCI and healthcare research, co-design has been widely used for acquiring knowledge from domain experts, such as healthcare providers, in the early stages of the design process [177–182]. To identify design opportunities for healthcare technologies, researchers have conducted co-design studies with chronically-ill teens and their parents [177], clinicians and their patients [178, 179], and older adults [181]. During the co-design workshops, researchers usually act as moderators to facilitate the design process, while stakeholders bring up the key design considerations and envision how they would use the design in real-world settings [176, 178]. This is achieved through mutual interactions with design widgets or prototypes that represent the design outcomes [176]. Often times, researchers employ paper-based widgets and sticky notes to encourage flexible creations and to make modifying the design on-the-fly easy [177, 182]. For example, to collect design requirements of a home care reminder system for older adults, Mcgee-Lennon and colleagues led several co-design workshops, during which they asked a group of older adult participants to interact with and modify a set of pre-designed prototypes [181]. Similarly, Kim and colleagues designed DataMD, a clinical dashboard for healthcare providers to record and review patients' data [179]. During the design stage, researchers worked with clinicians, healthcare informatics experts, electronic medical records (EMR) developers to come up with multiple design ideas, and then iterated on their ideas to finalize the interface design [179]. Such close collaboration among designers, researchers, target users, and other stakeholders can gather valuable insights that might otherwise be overlooked by either party, and thus drive innovative solutions [173, 175].

3.4.2 Evaluating Self-Tracking Technology

Evaluation is an important step in HCI research, during which researchers examine whether and how a proposed system solves problems, performs better than previous systems, brings positive experience to people's life, or informs design opportunities for future work [49, 183–185]. In evaluating self-tracking technologies, an important consideration is to situate people in a setting where they can capture and reflect on their own personal data. Therefore, instead of running lab studies that ask people to complete certain tasks in a constrained environment, researchers tend to conduct field studies to investigate the user experience "in the wild" [8, 11, 21, 90, 186–188]. While some of these studies focused on examining the systems' effects on behavior change, others focused on people's everyday interaction experience with the system or the quality of the collected data.

3.4.2.1 Eliciting Everyday Interaction Experience

As a first step to examining whether a system can be used as intended, evaluation is often tied with usability and feasibility [184, 185]. In the context of self-tracking, researchers commonly employed data collection studies (also known as diary studies) [8, 21, 186, 189, 190], in which the self-tracking system is considered as a research prototype to elicit lived interaction experience rather than a behavior change intervention. For example, Grimes and colleagues deployed EatWell in several households and asked family members to create shared nutrition-related stories; the study focused on analyzing different types of stories and understanding how participants' emotion and social relationships played part in their story-crafting experience [189]. In these studies, important evaluation metrics include participants' ***tracking adherence*** (e.g., number of data entries), ***data capture burden*** (e.g., time needed to track, perceived difficulty level), and their ***engagement*** with particular features built in the system (e.g., how often people revisit the collected data) [8, 11, 186, 189, 190]. Although such studies could not prove the long-term effect of the systems, they could help researchers understand the real-world user experience within a relatively short period, identify usability problems at an early stage of design, and unfold design opportunities for improving existing systems.

Echoed with Klasnja and colleagues' call on "*rethinking technology efficacy*," another way to examine the potential benefits of the proposed system is ***tying evaluation to behavior change strategies*** [49]. As Section 2.1.1.2 describes, self-tracking can lead to behavior change because of reactive effect, which occurs when people become aware of their behaviors and think about their behavior more frequently and seriously. In this light, many researchers start by looking into whether and how their self-tracking systems can ***raise awareness*** and ***enables reflection*** on the target behaviors [8, 11, 117, 190, 191]. For example, Choe and colleagues designed SleepTight, a mobile sleep tracking app that collects people's sleep-related data (e.g., bed time, sleep quality, coffee intake) [11]. From the usage log and debriefing interviews, the researchers revealed how participants access and reflect on their sleep data; they also identified different types of insights that participants drew from their sleep patterns, which could help them develop personalized strategies to improve their sleep habits [11]. In a similar vein, Rivera-Pelayo and colleagues built MoodMap, a web application that tracks workplace wellbeing including mood and energy level for individual workers and their team members [190]. The researchers qualitatively examined how the system enabled participants to reflect on the ways that their mood influenced their individual tasks and teamwork [190]. Through a four-week study ($N = 71$), the findings showed how increased self-awareness of emotional wellbeing helped participants better organize their workflow and increase cohesion within teams [190].

3.4.2.2 Assessing Data Quality

Given the assessment purposes of self-tracking (See Section 2.1.1.1), another commonly used approach to evaluate self-tracking technology is assessing the quality of collected data to examine whether the system can help people capture valid and useful information [10,186,192–195]. When assessing structured data that are automatically collected (e.g., physical activity captured by wearable devices), important factors determining data quality often include data accuracy, precision, and amount [192, 193]. But when it comes to assessing unstructured data that are manually collected by people, the standard of data quality varies depending on the assessment goals. In the context of dietary assessment, for example, good quality data should include sufficient information to meet healthcare providers' assessment needs, such as meal ingredients and portion size [195]. For individuals with specific health conditions (e.g., IBS patients), high quality data may require other details such as eating time and health symptoms that are specific and precise [10]. In a recent food journaling study aiming to examine how different input modalities affect data quality, researchers focused on the granularity and specificity of participants' self-reported food intake [101]. By grouping the food items into single food, decomposed single food, and aggregated food, the researchers also categorized the specificity level into generic, specific, and varietal to illustrate how people describe their food differently through written text, spoken language, or search results from existing food database [101].

While the rise of computer vision and machine learning techniques has made it possible for technology to perform data assessment, these techniques are still limited in generalizability (e.g., food recognizers that can only recognize a limited amount of food from photos [196]) and lack of tracking contexts (e.g., NLP services that can extract only food names but not other contextual data such as eating time [197, 198]). Therefore in many cases, researchers and healthcare providers still need to perform data assessment based on their professional experiences and judgements.

3.4.2.3 Examining Behavior Change

Because one of important goals of self-tracking is to achieve positive behavior change (See Section 2.1.1.2), previous work has prioritized behavior change as the main outcome of a self-tracking system [26, 199–201]. For example, Stawarz and colleagues compared different reminder mechanisms in habit formation by situating the research in the context of reporting daily lunch [201]. Through a four-week long study (N = 133, six study groups), researchers measured the average number of messages sent by each group as well as the changes in participants' habit strength (using existing questionnaires) [201]. In another study, Kim and colleagues developed TimeAware, a desktop widget to promote information workers' productivity with two different visual feedback displaying their productivity level [2]) [26]. Researchers conducted an eight-week long evaluation (N = 24, two study groups) [26], during which they first collected participants' productivity level for two weeks without showing any feedback (baseline period), and then introduced the two

[2] The productivity level was calculated by workers' computer usage, synchronized from RescueTime [129].

versions of feedback for four weeks (intervention period); lastly, they took away the feedback and kept tracking participants' productivity level for another two weeks (withdrawal period) [26]. By comparing the productivity change between the baseline and intervention periods, and between the intervention and withdrawal periods, researchers were able to figure out which design effectively improved participants' productivity [26].

However, behavior change is a long and complex process with high relapse rate [48]. To prove that a system has successfully led to behavior change, we need to conduct large-scale and long-term studies. However, such studies are rarely feasible in HCI research; instead, studies with HCI contributions mainly focus on evaluating early-stage prototypes and generating design implications [49]. Often times, even when behavior change is observed, it is still difficult to evaluate whether such change is sustainable [202]. Furthermore, emphasizing on behavior change may overlook the side effects or even lead to counterproductive outcomes resulting from the system [109, 203]. For instance, although tracking smoking activities seem to help lower the frequency of smoking, researchers found that revisiting one's smoking records sometimes could trigger the desire of smoking, making it even more difficult to quit [203]. Therefore, in addition to examining the behavioral outcomes of a self-tracking system, it is worth directing attention to people's everyday interaction experience—with both quantitative and qualitative approaches toward understanding *how* and *why* people perceive and use with the system in particular ways and *how* the systems influence their daily life [49] (See Section 3.4.2.1).

3.5 Chapter 3 Summary

In this chapter, I covered prior research related to the three self-tracking contexts that I examine in this dissertation: food practice, productivity, and exercise. I also described how HCI researchers design and evaluate self-tracking technology, along with rationales behind the evaluation methods.

Learning about the gap between the designs of existing food journaling tools and healthcare providers' assessment needs, I address the first research theme on multimodal self-tracking in the context of food journaling by asking **RQ1** (What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems?). Inspired by the co-design methods, I was motivated to conduct a co-design study with registered dietitians because they usually need to initiate food data collection from patients and utilize the data as part of their workflow [3].

To address the second research theme on integrating touch and speech input on mobile phones, the importance of collecting eating contexts motivated me to continue studying food journaling by asking **RQ2** (What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?) [4]. Learning from previous work on understanding personal productivity, I was inspired to decompose productivity into multiple dimensions that can be captured by different data types. Therefore, I ask **RQ3** (How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes?) and **RQ4** (How does the in-

[3] In Chapter 4, I describe a co-design study with dietitians to identify design opportunities for food trackers.
[4] In Chapter 5, I describe a speech-based food journaling study focusing on input burden and data richness.

put modality affect the data richness in unstructured data? in the context of productivity tracking) [5]. The study designs to answer **RQ2**–**RQ4** were based on how previous research elicited everyday interaction experience and assessed the quality of the collected data.

Lastly, having learned about the main challenges of performing consistent exercise and the promise of smart speakers' hands-free interaction to lower the interaction burden, I was motivated to examine the third research theme on smart speakers in the context of exercise tracking. Thus, I ask **RQ5** (How does a smart speaker complement and augment a mobile app in supporting consistent exercise?). The research design was based on how prior works examined behavior change and elicited everyday interaction experience [6].

[5] In Chapter 6, I explain how I designed and developed NoteWordy, a mobile app integrating touch and speech input to capture different types of data about their tasks and breaks.

[6] In Chapter 7, I explain how I designed and evaluated TandemTrack, a multimodal system coupling a mobile app and an Alexa skill on smart speakers.

Chapter 4: Co-Designing with Dietitians: Identifying Opportunities for Customizing Food Trackers

Chapter 4 forms the first step of this dissertation work. In this chapter, I describe a co-design study with registered dietitians to answer **RQ1**: What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems? Focusing on healthcare providers' perspective, the findings hold important implications for satisfying individual patients' tracking needs based on their health conditions and collecting relevant clinical data for better healthcare.

4.1 Introduction

Food tracking is a prevalent practice for individuals to gain awareness on their diet [204, 205]. From healthcare providers' perspective, the data collected through food tracking has important clinical values for assessing patients' nutrient intake and providing treatment [30, 206]. But these values only manifest when the data collected are relevant to individual patients' health concerns and providers' assessment needs. For example, irritable bowel syndrome (IBS) patients need to identify the food that triggers their IBS symptoms by tracking their food intake and symptom details [41, 108]; eating disorder patients need to capture their feelings about food, such as hunger/fullness level, to develop mindful eating habits [109, 110]. While individual patients' tracking needs vary, mainstream food tracking tools (e.g., MyFoodDiary [51], MyFitnessPal [52], Bitesanp [59]) predominately focus on collecting calorie and nutrients, providing little flexibility for people to choose what to track and how to track about their food. As a result, people appropriate other tools that are not designed for food tracking, such as spreadsheets [207], social media [113], and bullet journals [208].

The mismatch between the design of existing tracking tools and individuals' tracking needs suggests that food tracking tools should be customizable. Thus, my colleagues and I set out to understand the information that patients need to track from a dietitian's perspective, and rethink the design of food trackers to support their information needs. We prioritize dietitians' perspective than patients' because data collection in this context is usually initiated by the dietitians [40] and the data are being collected and utilized as part of dietitians' workflow. Specifically, we aim to identify what patients need to track to facilitate working with dietitians (**Tracking needs**), and how to customize food trackers to support patients' and dietitians' needs (**Tailoring tracker design**). In the following, I describe how we facilitated the co-design sessions, what we found, and what we learned.

4.2 Structuring The Co-Design Workshop

We conducted individual co-design workshops with six registered dietitians, started with pre-design activities to understand their typical workflow and to create their representative patient personas, and followed by debriefing interviews to understand participants' co-design experience. Because we were interested in identifying uniqueness and breadth of individual dietitians' practice, we conducted individual sessions instead of group sessions. This study was approved by the university's Institutional Review Board (IRB # 1132164-2).

4.2.1 Participants

To recruit registered dietitians, we first drew up a list of 68 dietitians whose contact information was found on the websites of various local nutrition services. Among the 68 dietitians who were contacted, six dietitians responded to us. They all met the following inclusion criteria: individuals who (1) are registered and accredited dietitians, (2) have been working as registered dietitians for more than 6 months, (3) have been providing services to patients with dietary problems, and (4) employ (technology or non-technology based) food diary in their practice. All six participants were female (all the dietitians in our recruiting pool were female), and their age ranged from 27 to 67 (M = 38.5, SD = 6.64). The six participants had diverse training from different regions within the U.S., and are currently working in different clinic environments (see Table 4.1 for their background [1]). According to participants' preferences, we conducted four co-design workshops at a research lab (P1, P2, P3, P5), and the other two at participants' office (P4, P6). Each workshop lasted from 70 to 90 minutes. Among the six workshops, three involved two researchers and the other three involved three researchers. At the end of the study, each participant was compensated with a $75 gift card.

[1] (P#) is an unique ID to denote each participant.

ID	Age	Gender	Practice years	Work environment	Training background	Expertise
P1	68	F	30	Private practice	Health Education & Nutrition Sciences	WM, ED, diabetes, GI
P2	34	F	11	Medical center & private practice	Mental Health & Dietetics	WM, ED
P3	27	F	2	Eating disorder treatment center & private practice	Public Health & Nutrition Sciences	WM, ED, GI
P4	43	F	20	Private practice & corporate wellness	Nutrition Sciences & Dietetics	WM, ED, diabetes, nutrition during pregnancy, digestive issues
P5	34	F	9	Eating disorder treatment center	Nutrition Sciences & Dietetics	ED, diabetes
P6	60	F	30	Private practice	Nutrition Sciences	WM, GI, ED, diabetes, pregnancy, rehabilitative, autoimmune, cardiac issues

Table 4.1: Participant profiles (WM: weight management, ED: eating disorder, GI: gastrointestinal distress).

4.2.2 Patient Persona

At the beginning of the study, we asked participants about their background, years of practice, and how they appply food diary in their everyday work. In particular, we asked each participant to describe two **patient personas** that they commonly see. Persona has been widely used in design as a tool to build empathy between users and designers [209]. In designing consumer health technologies, researchers found that patient persona was effective in addressing the needs and challenges of health consumers, especially for those who have comorbidities [210]. The patient personas participants described included the patient's age, gender, dietary problems, symptoms, and treatment goals (See Table 4.2 for detailed patient personas). The remaining conversations and co-design activities centered on these personas. Focusing on each patient persona, participants described their typical treatment workflow—for instance, what the first session looks like, what information to collect, and what tools to use for tracking and sharing data. Having them describe a concrete context of work and patient's health concerns, participants were able to reflect on their everyday practice so as to be prepared for the co-design activity. Serving as a warm-up, this activity also provided us a better understanding of participants' practices and the patients they commonly see.

4.2.3 Co-Design Activity

After the participants finished describing their patient personas, we used paper prototypes [211] as a tool to foster dietitians' creativity and to facilitate the co-design activity. Inspired by Kim and colleagues' survey on common tracking field types in commercial tracking applications [21], we provided participants with a set of paper-based widgets representing different data formats (e.g., text box, numeric, date, time, radio-button, location, Likert scale, image, checkbox). We provided images of external sources (e.g., Fitbit, Apple Watch, glucometer), assuming that data from these sources can be integrated, if needed. We also prepared blank widgets in case participants want to create a new field type on their own. The paper widgets were larger than the actual size of those shown in a mobile phone screen, which gave participants enough space to label, mark, and annotate. By assembling the widgets, participants could easily design the trackers and modify them. Figure 4.1 shows two examples on dietitian participants engaging in co-design activities.

Using a large white board (635 x 762 mm) as a frame, participants were asked to create one food tracker for each patient persona they described and to think aloud during

[1] I use "D-#" to denote diabetic patients, "WM-#" to denote weight management patients, "ED-#" to denote eating disorder patients, and "GI-#" to denote patients with gastrointestinal distress.

Patient ID[1]	Created by	Age	Gender	Symptoms & Health Conditions	Goals
D-1	P1	Mid-50s	M	Weight gaining, prediabetes (A1C = 7)	Not rely on insulin, maintain his job
WM-1	P1	30	F	Weight gaining, in good health	Identify what in her diet caused the weight gaining
WM-2	P2	11-16	F	Overweight, body-image focused, low self-esteem, anxiety	Build self-esteem, make food choices she feels good about, increase food variety
WM-3	P4	60	M	Overweight, new diabetes (A1C = 8.5)	Decrease calorie, balance glucose level
WM-4	P4	50	F	Overweight, in good health	Lose weight, decrease calorie, drink enough water
WM-5	P6	45-50	F	Overweight	Get healthier, lose weight
ED-1	P2	20	F	Anorexia Nervosa, over-restricting eating, over-exercise	Increase calorie & food variety
ED-2	P3	22	F	Anorexia & Orthorexia tendencies (non diagnosed)	Regain menstrual cycle, overcome social isolation & preoccupations on food
ED-3	P5	18	F	Other specified feeding or eating disorder (OSFED), Anorexia & Orthorexia tendencies, severe Obsessive-Compulsive Disorder	Improve life quality, overcome social isolation, increase calorie & food variety
ED-4	P5	45	F	Bulimia Nervosa, prediabetes, weight gaining, fatty liver	Decrease calorie, eat more protein
GI-1	P3	Mid-40s	F	Gastrointestinal (GI) distress, diarrhea, constipation	Identify the foods that trigger her GI symptoms
GI-2	P6	45	F	Gastrointestinal distress, sleep problem	Identify the foods that trigger her GI symptoms

Table 4.2: Patient personas that dietitian participants created during the co-design workshops (D: diabetes, WM: weight management, ED: eating disorder, GI: gastrointestinal distress).

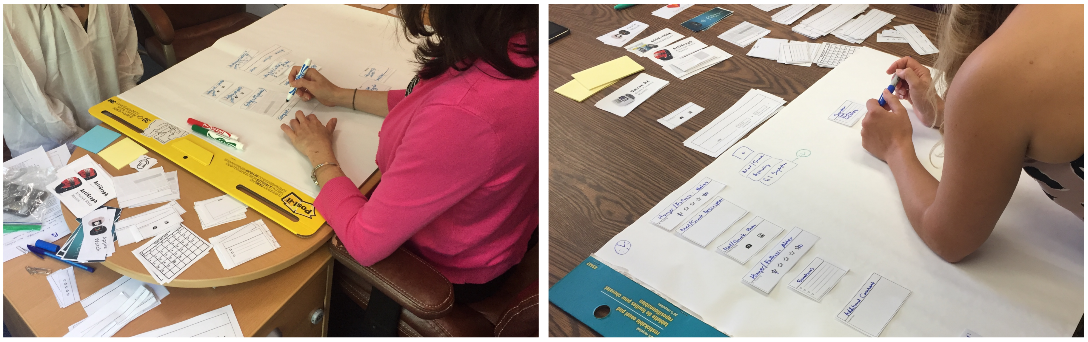


Figure 4.1: Dietitian participants designing food trackers with researchers in their office (left) and a research lab (right) using the paper widgets we provided to facilitate the tracker customization.

the process. Participants could choose any widgets they liked to use, modify existing widgets, and annotate the items they added. To follow up with their tracker design rationale, we asked questions, including why such information is important to track, whether this is required or optional, why they choose to use a particular widget, and how often this information is needed to be tracked. When multiple widgets were added into a food tracker, we asked the participant to think about how they want to arrange the widgets regarding tracking order and priorities.

Some participants dived into the co-design activity right after we explained the procedure, whereas others showed hesitation due to the unfamiliarity with this process and worried about the quality of their design. To reduce their concerns, we clarified that our goal was to understand the tracking needs of different patients instead of evaluating their design. We gave participants as much time as they needed, and prompted them to start with the most important information they would like to have by asking "*what information do you currently collect from the patient*," "*how do you currently collect this information*," and ‘ ‘*what is the information that you wish to collect but couldn't*."

4.2.4 De-briefing Interviews

At the end of the study, we asked participants to reflect on their tracker design, including how they would use the data collected from the food trackers they designed and how they would want to share the data (synchronously, asynchronously) with the patient. Participants also reflected on the experience of participating in the co-design workshop.

4.2.5 Data Analysis

The dataset includes the audio recordings of the entire co-design workshops and paper prototypes that participants created. All the audio recordings were transcribed to text. Three researchers collaboratively analyzed the data: we first analyzed two audio transcriptions individually to note prominent themes using open coding; then we met several times to discuss each theme to generate affinity notes and to update affinity diagram [212]. Next, we worked with one of the researchers to repeat this process in analyzing the rest of the transcripts. Last, we digitized the paper prototypes using Sketch [213], and analyzed the tracker design by referring to the audio transcript section where participants described the rationale behind their tracker design. We specifically examined tracking items, tracking frequency, and data format.

Despite the relatively small sample size, the study generated rich data, including 12 patient personas and 12 paper-based food trackers, and six workshop audio files which were transcribed to 58,858 words. The data collected during the study enabled me to uncover the commonalities as well as differences among the patients and their tracking needs, and multiple customization dimensions for food tracker design.

4.3 Results

Participants were engaged in the co-design workshops and excited about the idea of creating customized food trackers. During the pre-design activities, they reported treating diverse patients, whose ages range from 11 to 85, and who are mostly female (75–90%). Based on the patients they commonly see, participants described 12 patient personas, which we categorized into four groups based on their primary dietary problems: diabetes (D), weight management (WM), eating disorder (ED), and gastrointestinal distress (GI). These personas shared some characteristics regarding age, dietary problems, symptoms, and goals (Table 4.2), but they were also different in unique ways. In the following, we first provide background information on a typical treatment workflow and current ways of using food diary data in participants' practice. Then we describe their diverse information needs regarding what to track (tracking needs) and how to track (tailoring tracker design).

4.3.1 Current Treatment Workflow and Tracking Tools

All participants consider food tracking an important part of the treatment. When a new patient visits, three participants (P1, P4, P6) require them to bring a food diary, while others (P2, P3, P5) introduce the food diary during the first or second session based on how the meeting progresses. Patients are asked to continue keeping a record of food diary throughout the treatment. Participants review the food diary data during the in-person session and between visits to provide feedback, discuss patients' progress, and help them troubleshoot. From patients' food diary, participants *"look for food patterns"* (P1) and *"identify potential problems"* (P6). During the face-to-face meeting, participants examine other factors that might affect patients' diet such as medical history and social environment. By synthesizing the information from different sources, participants provide education, help patients set diet-related goals, and customize meal plans.

Although the high-level treatment workflows are similar across our participants, the specifics of how each participant practices vary. For instance, while most participants meet their patients in the clinic, P2 goes to her patients' home to help them set up the cooking environment. The length of each session also greatly varies depending on participants: the first session could last from 15 to 150 minutes, and later sessions could last from 15 to 45 minutes. Their ways of working with patients, motivation strategies, and the pace of treatment differ from dietitian to dietitian. Moreover, the same dietitian could practice differently depending on the type of patient they see.

As for the food tracking tools they currently employ, we saw a mix of paper-based diaries and mobile apps (i.e., MyFitnessPal (P1, P4), Fitbit (P1), Recovery Record (P2, P3, P5), Healthie (P3), Lose It! (P4, P6), Cronometer (P4)). Some participants also use 24-hour food recall (P5, P6), spreadsheet (P4), and email (P3). In most cases, participants recommend their patients to use any tool that suits individuals' preferences (e.g., paper-based diary for older patients and mobile apps for younger patients). We did not find any participants who currently customize tracking items for individual patients; however, for eating disorder patients, a specialized tool designed for this population (i.e., Recovery Record [214]) was often recommended.

4.3.2 Tracking Needs

By *tracking needs*, we mean the data that can only be captured through patient's tracking to fulfill dietitians' information needs. We identified 32 unique items that can be collected from patients' food diary to aid dietitians in their treatment. These items were grouped into five categories: *food* (7), *reflection* (12), *activity* (6), *symptom* (4), and *physical state* (3) (See Table 4.3 for details). Depending on patients' dietary problems and dietitians' practice, the necessity and importance of these tracking items vary. Tracking needs consist of not only the factual information (food, activity, symptom, physical state) for dietitians to identify patterns of behavior, but also subjective data (reflection) for patients to contemplate their own eating behaviors. In this regard, food tracking served a dual purpose of assessment and treatment [2].

4.3.2.1 Food

Tracking food-related information was expected for all of the patient personas. Specifically, all of them were expected to capture meal time (start/end) and meal type (breakfast, lunch, dinner, snack), from which dietitians can infer regular/irregular diet, job situation (e.g., on a shift), or major life transition. Items in the food category include food and its contextual information such as location: P4 and P5 wanted to know the location where patients eat, because they were interested in whether the meals were homemade or store-bought (e.g., prepackaged or from restaurant).

We also found differences in tracking needs for patients with different dietary problems. For weight management patients, especially those having a goal of reducing calorie

ID	Created by	Food	Reflection	Activity	Symptoms	Physical states
D-1	P1	food items, meal type, time, portion size.		sleep		glucose, BP
WM-1	P1	food items, meal type, time, nutrition facts, portion size				weight
WM-2	P2	food items, meal type, time	body image, things to be proud of, self-care behaviors, treats, food groups, emotion on food	exercise (type, location)		
WM-3	P4	food items, meal type, time, nutrition facts, portion size, location				glucose
WM-4	P4	food items, meal type, time, nutrition facts, portion size, location, water	hunger/fullness level, eating strategy			
WM-5	P6	food items, meal type, time, nutrition facts, portion size, water	mood, hunger satisfaction rating	exercise (type, time, duration, intensity), sleep		weight
ED-1	P2	food items, meal type, time	body image, things to be proud of, self-care behaviors, challenge food, emotion on food	exercise (type, duration)	ED-behavior	
ED-2	P3	food items, meal type, time	hunger/fullness level, mood	exercise (type, time)	ED-behavior	
ED-3	P5	food items, meal type, time, location	hunger/fullness level, mood, thoughts		ED-behavior	glucose, weight
ED-4	P5	food items, meal type, time, location	hunger/fullness level, mood, thoughts		ED-behavior	
GI-1	P3	food items, meal type, time	hunger/fullness level, mood	exercise (type, time)	GI symptoms, time	
GI-2	P6	food items, meal type, time, nutrition facts, portion size	mood	exercise (type, time, duration, intensity), sleep	GI symptoms, time, severity	

Table 4.3: Items that can be captured by patients with dietary problems to facilitate collaboration with dietitians. A total of thirty two tracking items were identified, and then grouped into five categories (BP: blood pressure).

intake (WM-1, WM-3, WM-4, WM-5), it was recommended to track nutrition facts (e.g., calorie, carbohydrate, fat, sugar, sodium, fiber) and portion size. Through tracking these numbers, patients could learn how to figure out *"the value of their foods"* (P6), compare different foods (P4), and try to *"balance calorie in and out"* (P1). Despite having a weight management issue, WM-2 was *not* recommended to track nutrition facts and portion size because of her low-self esteem issue. For patients with eating disorders (ED-1, ED-2, ED-3, ED-4) or with low-self esteem (WM-2), tracking nutrition facts and portion size can be counterproductive, because patients are easy to get *"obsessed"* (P2), and the numbers can be *"overwhelming"* (P2, P3) and even *"trigger ED-behaviors"* (P5).

4.3.2.2 Reflection

To develop awareness and mindfulness, participants suggested that patients reflect on their food choices, their body, activities, and feelings. As some dietary problems are highly related to mental health issues (e.g., body image focused, low self-esteem), some dietitians used tracking as an intervention to foster self-reflection and mindfulness [215].

Participants suggested a variety of items to reflect on, with some overlaps across different patient types. Three dietitian participants (P3, P4, P5) employed a standard measure—the hunger/fullness level for patients with different dietary conditions (WM-4, ED-2, ED-3, ED-4, GI-1). The goal of tracking hunger/fullness level was to help patients build trust in their internal body cues, such that they can eventually *"make independent food choices"* instead of being affected by external cues such as *"diet magazines and nutrition labels"* (P3). P6 was similarly interested in capturing internal body cues, but with a different measure—hunger satisfaction rating, which has different scale and interpretation from hunger/fullness level (e.g., a person can feel full but not satisfied). Besides, considering that patients' mood can interplay with the food they eat, three participants (P3, P5, P6) suggested mood tracking for their patients (WM-5, ED-2, ED-3, ED-4, GI-1, GI-2). To motivate patients to form a habit of reflecting on *"what's going through their head and body,"* P5 also wanted ED-3 and ED-4 to track any thoughts they have and anything they like to express.

Different from other dietitians, P2 was particularly keen on reinforcing positive thinking for her patient personas (WM-2, ED-1). P2 encouraged them to reflect on their body image, things to be proud of, and activities conducted to "honor your body" (e.g., self-care behaviors: dancing, taking a bath, going for a walk, talking with friends). She also emphasized the importance of tracking emotion towards food, which intends to help patients make food choices that make them feel good.

Participants also recommended individually-tailored reflection topics. For example, pointing out WM-2's mental health issue, P2 suggested her having treats as a praise of making progress, and reflecting on the food groups to make sure she is *"getting all the different food groups."* Given that eating disorder patients often have certain "challenge foods" (i.e., the food they are afraid of eating), P2 suggested ED-1 to reflect on her challenge foods to overcome such fear. To keep WM-4 mindful of the food portion size, P4 wanted her to reflect on her eating strategy (e.g., "was I thinking about eating half of it?").

4.3.2.3 Activity

Exercise and sleep were two activity types that were brought up during the co-design. Opinions were divided on whether to track exercise. Three participants (P2, P3, P6) were keen on exercise tracking for the personas they created. P3 wanted to see exercise type and duration for both personas (ED-2, GI-1); and P6 wanted to see more details (e.g., intensity) to understand how patients spend their energy and how their exercise might relate to their food practices for both personas (WM-5, GI-2). In addition, P2 emphasized that the purpose of having eating disorder patients track exercise is to prevent extreme exercise while encouraging light exercise.

However, not all participants were in favor of tracking exercise, as P4 explained: *"people subtract the calories [consumed from exercise], [...] And if this is not accurate, they're eating more calories, and then they're not losing weight"*.

Two participants (P1, P6) were interested in tracking patients' sleep data. P1 recommended D-1 to track sleep because she believed that diabetes and sleep problems are closely related. P6 suggested both WM-5 and GI-2 track sleep because sleep can affect their diet, for example: *"You don't sleep, it changes what you want to eat the next day, you want fat and sugar."* (P6).

4.3.2.4 Symptom

ED and GI patients experience specific symptoms, which need to be tracked. Tracking symptoms can help dietitians find out the source of problem and provide appropriate treatment and support. When treating GI patients, participants (P3, P6) wanted their symptom information to include detailed descriptions (e.g., diarrhea, constipation, gas) and time stamps to identify the foods that trigger their GI symptoms. In addition, P6 mentioned that capturing the severity of the symptom is also helpful.

All participants (P2, P3, P5), who created ED personas, stated that they need to know ED-behaviors (e.g, purging, over-exercising, vomiting, and use of laxatives), which are considered symptoms, and thus to be tracked. Being aware of ED symptoms allows participants to provide support when needed, while enabling patients to understand how their ED-behaviors occur and learn to cope with them: "*It's sort of like separates this from being an automatic behavior, if they kind of spend some time thinking about like, okay like, I ate food, I did what I was supposed to, it created these urges, [...], It might also bring out like urges to purge.*" (P3).

4.3.2.5 Physical State

Physical states such as weight, blood glucose, and blood pressure were of interest to some participants (P1, P4, P5, P6). For patients with diabetes (D-1, WM-3, ED-3), tracking blood glucose level was necessary; and for some WM and ED patients (WM-1, WM-5, ED-3), tracking weight was expected. Capturing physical states could help participants examine what types of food or activity might cause changes in these health indicators. However, for ED-3, P5 suggested that the weight information would not be captured at the clinic, because it can backfire her ED-behaviors, which we reported in more details in section 4.3.3.4.

4.3.3 Tailoring Tracker Design

In the above section, we reported how tracking needs might differ depending on patients' dietary problems and dietitians' style of practice. These differences on tracking needs were manifested in the tracker design. Besides customizing what items to track, participants also tailored the trackers by incorporating when to track (timing/frequency of tracking), how to track (data format), how to support tracking, and what to share between dietitians and patients.

4.3.3.1 Timing and Frequency of Tracking

Participants expected patients to track different items at various time- and frequency resolution—for instance, tracking with food, when an activity or symptom occurs, once a day, twice a day, or once a week. In the case of *food*, participants expected patients to track their food whenever they eat, right before or after they eat, and together with their food-related *reflection* (e.g., hunger/fullness level before and after each meal, mood before each meal, emotion on food after each meal) (P2, P3, P4, P5). Other types of reflection may be tracked less frequently, such as once a day (e.g., body image, self-care behaviors) or once a week (e.g., treats, food group covered) (P2). *Symptoms*, with their exact time stamps, needed to be tracked whenever they occur (P2, P3, P5, P6). *Activity* and *physical states* were expected to be tracked on a regular basis, such as once a day (e.g., exercise, sleep, blood glucose, blood pressure) (P1, P2), twice a day (e.g., blood glucose) (P4), or once a week (e.g., weight) (P1, P6).

Tracking item	Data format (#)
food item	text & audio (4), text & audio & photo (8)
meal type	text (3), drop-down menu (9)
meal time	auto-generated time (4), drop-down menu (8)
meal location	auto-tracked location (12)
portion size	text (2), photo (2), drop-down menu (2)
nutrition fact	auto-generated text (2), barcode (3)
water	drop-down menu (1), add button (1)
glucose	Glucometer (3)
BP	BP monitor (1)
weight	clinical scale (2)
mood	checklist (2), audio & checklist (2), emoji (1)

Tracking item	Data format (#)
hunger/fullness level	Likert scale (5)
hunger satisfaction rating	Likert scale (1)
eating strategy	text (1)
body image	text (2)
things to be proud of	text (2)
self-care behavior	checklist (2)
emotion on food	text (2)
challenge food	text (1)
food group	checklist (1)
treats	text (1)
thoughts	text (2)

Tracking item	Data format (#)
exercise type	text (4), checklist (2)
exercise time	Fitbit (2), auto-generated time (2)
exercise location	auto-generated location (1)
exercise duration	text (1), text & Fitbit (2)
exercise intensity	text & Fitbit (2)
sleep	rating & Fitbit (1), Fitbit (2)
ED-behavior	text (1), checklist (3)
GI-symptom	checklist (2)
symptom time	auto-generated time (2)
symptom severity	Likert scale (1)

Food Physical state Reflection Activity Symptom

Table 4.4: Data format that dietitian participants expressed to capture different tracking items.

4.3.3.2 Data Format

As participants assembled paper widgets of different field types, they devised how to best capture tracking items in which data format (Table 4.4). In addition to the widgets we provided, participants created two new widgets: a barcode scanning and emoji. Of the 32 tracking items, 19 were formatted using one type of widget (e.g, GI symptoms: checklist), ten were formatted using two types of widgets (e.g., exercise: text box and Fitbit data), and three were formatted using three types of widgets (e.g, food item: photo, audio, text box). Text box and checklist were most commonly used.

Figure 4.2 shows the digitized version of the paper-based prototype for WM-4, WM-5, ED-1, and ED-3, designed by P4, P6, P2, and P5 respectively. The same item can be tracked in different formats. For example, food items can be tracked by text, audio, photo, or a combination of these input modality. Portion size of food can be tracked with a text box (WM-5, Figure 4.2-b), drop-down menu, or before/after meal photos (WM-4, Figure 4.2-a). When tracking water intake, a drop-down menu (Figure 4.2-a) was used to capture total daily water intake, and a counter (Figure 4.2-b) was used to capture in-situ water intake. For mood tracking, P6 used a checklist of emoji (Figure 4.2-b), while P5 provided an option of speech input via audio recording (Figure 4.2-d). P5 pointed out that for eating disorder patients, speech input can afford them to record frank thoughts without feeling *"shame about the things they logged,"* because there is no visual feedback. P2 also agreed that speech input can be helpful for patients to "*track more detailed eating experience and freely express how they feel*." Furthermore, P3 and P6 highlighted the potential benefits of speech input to lower the data capture burden: "*Voice would work, because for me it's like anything that makes the logging easier and reduces barriers, so that they just log [their food] consistently could be cool.*" (P3). Another example of capturing the same information in different formats is exercise data: P6 chose to synchronize the automatically captured data from an external source (i.e., Fitbit) for WM-5, but P2 chose a text box for ED-1 because she concerned about the patient being "*obsessed over the devices.*" In addition to using different widgets, P6 suggested that when tracking the severity of GI symptoms, the scale can be flexible: "*Some people when I tell them 1 to 10 they go 12. Okay. We don't have a 12, but they're letting me know that's how severe it is.*"

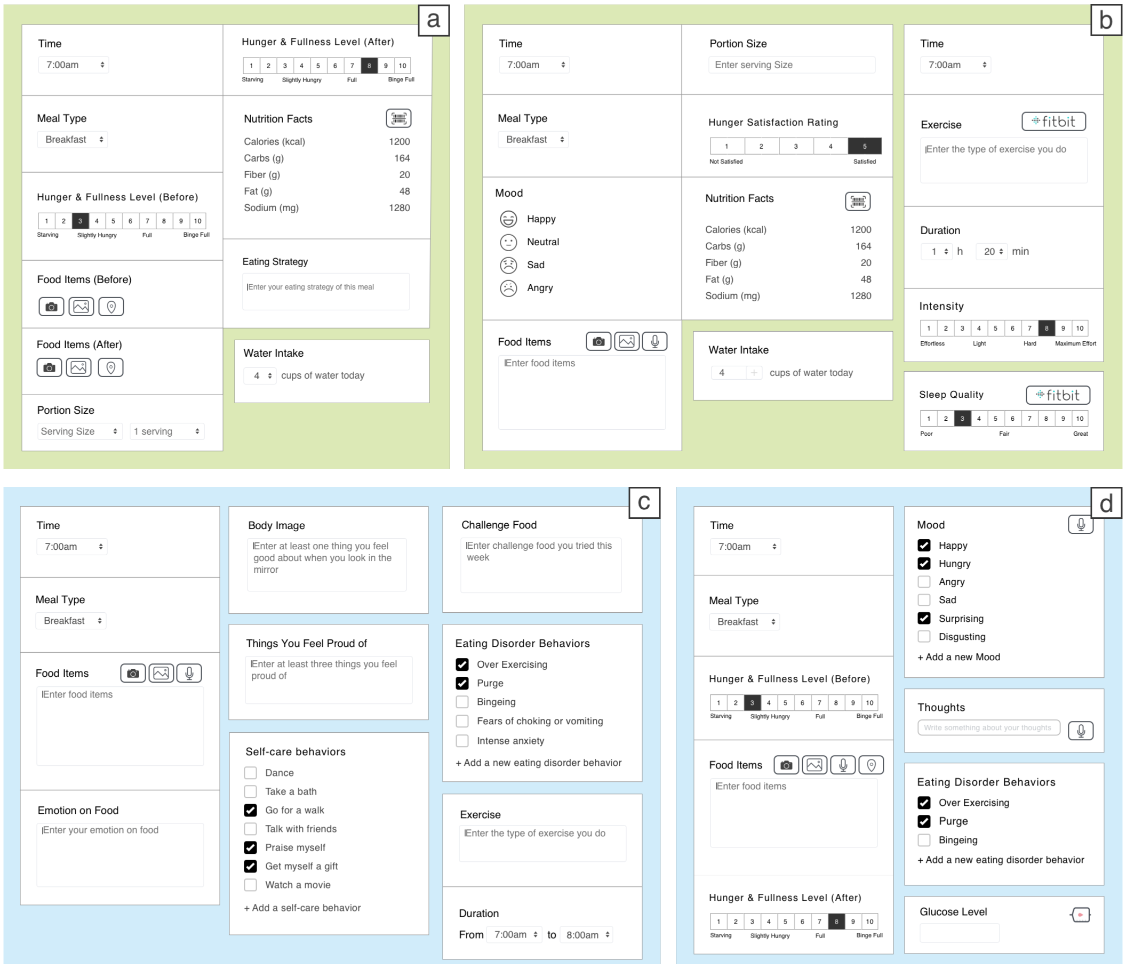


Figure 4.2: The digitized version of paper-based food trackers for WM-4 (a), WM-5 (b), ED-1 (c), and ED-3 (d). Items grouped together are meant to be tracked together at the same time. Icons next to the title represent alternative ways to capture the information (e.g., taking a photo is an alternative way to capture food items).

4.3.3.3 Supporting Features

Although the focus of the co-design activity was to identify tracking needs, participants naturally expanded the scope of design to devise ways to support patients in general. For example, P1 and P4 designed reminders to encourage patients to drink water (D-1, WM-3, WM-4), eat snack (D-1), and watch calorie limit (WM-1). P5 designed a prompt that automatically notifies ED-3 and ED-4 (as a positive reinforcement) when they had a "challenge food." To provide support between visits, P2 designed a help button for emergency contact; P3 and P5 wanted to access patients' data as they come in and make comments (ED-2, GI-1, ED-3, ED-4); P6 recommended a chat room where she can talk to WM-5 and GI-2 through instant messages. Furthermore, P5 and P6 wished that when patients log any negative mood, the tracker can be smart enough to provide in-situ support using external sources such as a list of coping skills for anxiety management (e.g., *"meditation, gratitude"*) and links to educational resources (e.g., *"body positive books and podcasts."*).

4.3.3.4 Data Sharing Preferences

While most of the tracking data were expected to be shared between patients and dietitians, it was not always the case. Depending on the sensitivity of the information and patients' acceptance, some items were more appropriate to be left with patients only, and others with dietitians only. For example, because eating disorder patients *"value low weight and [tend to] restrict food,"* P3 and P5 preferred not to have them track their weight. Instead, they record patients' weight information every time when a patient visits a clinic. On the other hand, although information such as personal thoughts and emotion is helpful for patients to track, some patients might not want to share with dietitians due to *"the feeling of shame and fear of judegment"*(P5). As such, P5 suggested that patients decide what to share with providers. As the treatment progresses, patients may be *"willing to share more with the clinicians"* because shame might have decreased throughout the recovery process (P5).

4.4 Implications

This study extends previous personal informatics research in two regards: First, we identify various customization dimensions in food tracking based on the similarities and differences across patients' condition and dietitians' style of practice. Second, the way we structured the co-design workshops provides insights for others interested in working with healthcare providers to identify design opportunities. In this section, I discuss implications for designing food trackers tailoring to patients and dietitians' needs and opportunities for leveraging multimodal input.

4.4.1 Customizing Trackers to Generate Relevant Data With Multimodal Data Input

We were motivated to conduct this study to address limitations in current tracking tools, one of which is the lack of customizability in the tool design [3]. Such limitation makes it difficult to capture relevant data for stakeholders. In the healthcare field, the inability to customize tracking items frustrates patients and providers, hindering them from effectively utilizing patient-generated data (PGD) [40,41]. To generate clinically relevant data from patient's self-tracking, Zhu and colleagues suggest involving clinicians early, preferably during the tracking configuration stage so that clinicians can provide concrete guidance on what to track and how to track [40]. Furthermore, to design self tracking tools to generate clinically-relevant data, Choe and colleagues suggest HCI researchers working closely with clinical stakeholders at the early phase of design stage [216].

Through co-designing with dietitians, we identified a set of customization dimensions, including tracking items, timing and frequency of tracking, data format, among others. These can can be captured automatically with wearable or medical devices, or manually with touch, photo, or speech input (audio recordings). When given a chance, participants customized food trackers based on the health conditions patients experience, as well as their style of practice. Although the former is well exemplified in various tools designed for specific patient groups (e.g., Recovery Record for eating disorder patients), the latter has been less explored. When it comes to deciding which data format to use, dietitian participants not only considered their own information needs but also how easy it is for patients to consistently track the data and how the input modality affect patients' willingness to frankly share their thoughts. In particular, their rationale of using speech input to capture more detailed information and lower the data capture burden suggested the potential for unconventional input modalities (e.g., speech, video) to better support people's tracking experience.

In this work, we were surprised to observe the diversity of treatment style, while identifying commonalities across dietitians. As such, we see opportunities in supporting dietitians to create and share "tracking templates" for different patients, which can be uploaded for other dietitians to search, download, adopt, and modify. Dietitians can pick different tracking templates for different patients created by either themselves or by other dietitians and fine-tune the template for each patient. This approach can reduce efforts required to configure trackers from scratch whilst satisfying individuals' tracking needs. Although OmniTrack, an open-source customizable tracker, allows people to customize trackers for their respective tracking needs [21], it does not support the creation and sharing of tracking templates. To strengthen the reusability of customizable tools, studies like this can inform the design of tracking templates for dietitians as well as patients having different dietary problems. Furthermore, we envision that such design approach can be applicable to other clinical domains beyond food tracking to make the self-tracking data relevant to clinical contexts.

4.4.2 Supporting Patient-Provider Collaboration

Collecting and sharing patient-generate data is a collaborative work in which a provider and patient play an equally important role. Although we identified various customization dimensions with multimodal input to create food trackers from providers' perspective, the findings are limited without soliciting patients' perspective, as they too have the needs to customize tracking items [11,217]. While supporting providers to customize tracking dimensions, future work should also enable patients to add personally relevant and meaningful items to the tracker. Thus, involving patients during the design and evaluation stage is an important next step.

In addition, our study indicates that self-tracking in the clinical context is a dynamic process. The tracking needs could change as the treatment progresses, which suggests that tracking tools should support providers to revise the tracking regimens based on the stage of treatment. In the meantime, patients' data sharing preferences may change depending on their recovery progress and relationships with providers. Therefore, the tracking tool should also enable patients to adjust what to share, whom they share with, and when to share. Going forward, it warrants real-world deployment studies to examine how such customizable trackers affect the collaboration between patients and providers.

4.4.3 Using Patient Persona for Contextualization

In our work, the process of creating and sharing personas helped both researchers and participants be contextualized in patients' experience before starting the hands-on design activities. Patient personas allowed participants to articulate their design precisely and realistically. When we asked participants why they decided to track specific information or use a particular widget, their answers were closely tied to the patient persona they were designing for. As participants added more tracking items, they constantly thought about whether the information is necessary, which widget to use, and whether the information is appropriate to share.

Typically, significant user research precedes persona creation; but in our case, each participant could easily describe two patient personas based on years of clinical experience. The descriptions were detailed and nuanced, although it was inevitable that personas reflected the perspective of providers more so than that of patients. It may be presumptuous to think that these patient personas perfectly capture patients' lived experience and their concerns. However, given that the goal of this work was to understand how patients' tracking can facilitate working with healthcare providers and fulfill their information needs, we believe that integrating patient personas in the co-design session was a good first step to bridge the information gap.

4.4.4 Fostering Creativity Through Paper-Based Widgets

We believe that the paper-based widgets were critical in fostering dietitians' creativity in the design process. Before the co-design activity, while being asked about what information each patient persona needs to track, participants' answers were mostly constrained by the tracking tools that they are currently using (e.g., MyFitnessPal, Recovery Record, paper-based diaries). After we introduced the paper-based widgets, however, participants started to think about more possibilities: besides what the current tools capture, they considered whether those tools are capturing the necessary metrics appropriately, what else they would need for providing better treatment, and what patients would need for their reflection. Drawn from previous work that used modularized data fields for tracker customization [21], we created a paper version of the modularized data fields for the co-design activity. In addition to the traditional text- and form-based data capture, we also wanted to give options for different data capture modalities, by creating paper-based icons for microphone (to signify audio recording), camera (for photo-based capture), and map marker (for location capture). The widgets provided in the form of modularized data fields served as building blocks for participants to start the design process with ease. However, we believe that it is important to provide opportunities to think beyond what we provide, such as by providing blank notes and encouraging to annotate the widgets.

4.5 Chapter 4 Summary

In this chapter, I report findings from six individual co-design workshops with registered dietitians. During the workshops, dietitian participants created representative patient personas and designed food trackers for each persona. The findings suggested how customizing food trackers composed of multimodal data input can fulfill dietitians' information needs—the wide range of potential tracking items with their timing and format of tracking could potentially generate clinically meaningful self-tracking data. Incorporating patient personas and paper-based widgets helped us working effectively with healthcare providers and solicit concrete design ideas. This work calls for a new type of customizable tracker that supports patients and providers to collaborate around data tracking and sharing. In addition, the ways that dietitians applied different input modalities (e.g., text, photo, speech) highlighted the opportunities for building multimodal self-tracking systems, which provide multiple input options to support personal data collection based on people's tracking goals and personal preferences.

Chapter 5: Understanding How Speech Input Supports Food Journaling

In Chapter 4, I addressed the first research theme of this dissertation on identifying design opportunities for multimodal self-tracking. In this chapter, I aim to address the second research theme (integrating speech and touch input on mobile phones to support self-tracking) by answering **RQ2**: What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?

5.1 Introduction

Food journaling supports a variety of health goals such as weight loss and balanced diet [12]. In the digital era, we see numerous technologies that support food journaling, including photo [118], barcode scanning [119], accelerated search [120], and smart sensors [122]. While these technologies predominantly focus on capturing calories and nutrients, researchers have highlighted the importance of capturing relevant factors that play parts in people's food practice (e.g., time of eating, mood, eating environments), which are essential for individuals to perform self-reflection [90, 107] and for health professionals to make personalized diet recommendations [10, 218].

Because food practice is highly individualized, it is difficult to capture "unified" key factors that influence everyone's food practice with automated approaches [218, 219].

Therefore, prior research often employed questionnaires and interviews to identify factors that influence food decisions in different contexts such as home, restaurants, and grocery stores [220–223]. However, these approaches heavily relied on people's retrospective memory, rather than examining how people make their food decisions *in-situ*.

In this light, My collegues and I created FoodScrap, a mobile app to capture individuals' food components, preparation methods, and food decisions in free forms. To understand the experience of capturing everyday food practice using speech input, we conducted a one-week data collection study deploying FoodScrap to 11 participants from diverse food cultures. We also measured participants' perceived data capture burden using a set of subscales from User Burden Scale (UBS) [224] and conducted debriefing interviews. Focusing on data richness (i.e., the amount of data and the level of details) and data capture burden (i.e., how easy or difficult to capture data), the following describes how we created FoodScrap, what we found from the data collection study, and implications for incorporating speech input into self-tracking tools.

5.2 FoodScrap

FoodScrap was created with OmniTrack for Research, a web-based research tool that enables the creation and deployment of a flexible mobile self-tracking app [225]. Because this study focuses on examining the data richness and data capture burden of speech input, FoodScrap was designed as a data collection instrument, which does not provide detailed feedback except for the recorded log entries.

5.2.1 Journal Design

FoodScrap consists of three food journals: Main Meal Journal, Snack Journal, and Skip Journal. Figure 5.1 illustrates the interface of Main Meal Journal. All questions included in the journals were required. The logging time and session timestamps were automatically captured. Details of the mobile app UI are described in [21]. The Main Meal Journal captures the following information for each meal:

Q1. The type of the meal: breakfast, lunch, dinner, and brunch (as an alternative for breakfast or lunch)

Q2. Eating duration (start and end time)

Q3. A photo of the meal

Q4. *"Please describe the meal components and preparation methods."*

Q5. *"Why did you eat at this time rather than earlier or later?"*

Q6. *"Why did you choose this food instead of other options?"*

Q7. *"When did you make the decision to eat this food?"*

Q8. *"Why did you eat this much food?"*

We asked people to take a food photo in Q3 so that they can remember to log their meals later. To ensure that people capture their meals close to the time they eat, Q2 only takes a time range that falls within the current day. In particular, we designed four guided prompts asking why people decide *when to eat*, *what to eat*, *how much to eat*, and *when they make the decision* (Q5 to Q8 (Figure 5.1), which are key questions in examining the multifaceted aspects in food decision-making [219, 226]. Although understanding how

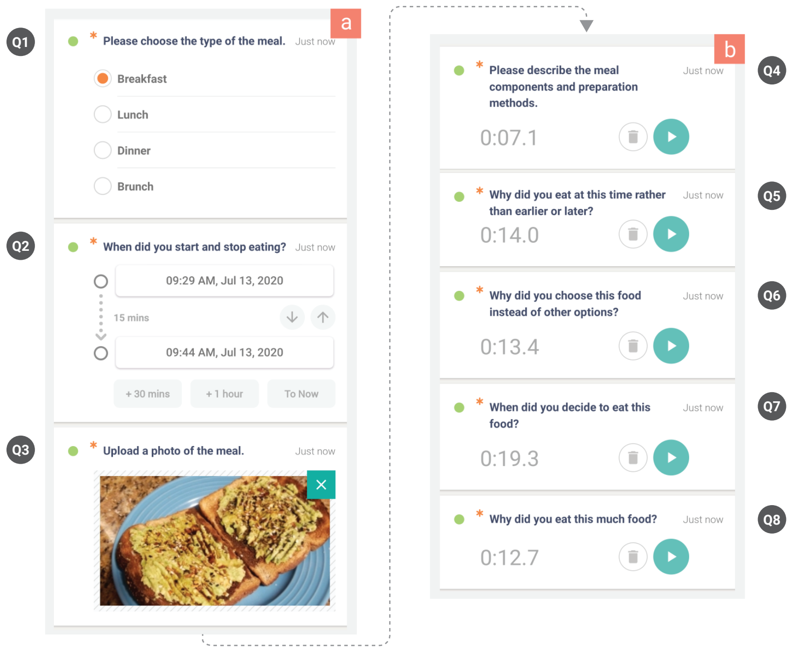


Figure 5.1: The data capture screen of Main Meal Journal in FoodScrap: (a) questions on meal type, eating duration, and photo of the meal; (b) questions on meal components, preparation methods, and food decisions.

people make food decisions has long been an interest in food science research, a majority of prior works employed questionnaires and interviews to retrospectively identify factors that influence food decisions [220–223] rather than examine how people make their food decisions *in-situ*. These questions take free-form audio recordings as responses, providing the flexibility for people to express additional thoughts.

Snack Journal asks the same information as Main Meal Journal, except for Q1 (meal type). We also created a Skip Journal to capture the main meals that people skip (excluding snacks) with three questions: the type of the meal that was skipped (SK1); *"When did you decide to skip the meal?"* (SK2); and *"Why did you decide to skip the meal?"* (SK3).

FoodScrap follows the design of commonly used voice recording interfaces (e.g., Samsung Voice Recorder [227]), allowing people to pause and resume recording. When recording is complete, people can play back their recording or delete the recording to start over. To exclude the effect of speech recognition errors that might influence user experience [228], we did not provide dictation or transcription support for speech input.

5.2.2 Daily Reminders

To capture as many journal entries as possible, we set up reminders for all three main meals (i.e., breakfast, lunch, dinner), and an additional summary reminder at the end of the day. We personalized the reminder times based on each participant' estimated eating time. The end-of-day reminder was set to be sent one hour after the dinner reminder. To reduce interruption, each reminder was triggered only when the participant had not logged their meals by the reminder time. For example, if a participant had captured their lunch before their lunch time, they would not receive a lunch reminder. If a participant had captured all their meals before the end-of-day reminder time, they would not receive the end-of-day reminder. Journal entries were considered valid as long as they were submitted within the same day, thus FoodScrap prevents people from submitting expired entries.

5.3 Methods

We deployed FoodScrap for seven consecutive days and conducted a post-study survey and debriefing interviews. Due to the COVID-19 outbreak, we interacted with participants remotely via a Zoom video call [229] (in June-July 2020). The study was approved by the university's Institutional Review Board (IRB # 1132164-14). Unlike traditional self-tracking studies that focused on examining how tracking tools influenced participants' tracking adherence [11] and behaviors [45], this work instead aimed at analyzing and understanding the nature of the captured information. Therefore, we structured our compensation to minimize missing journal entries without influencing the amount of data captured, which we describe in subsection 5.3.2.

5.3.1 Participants

We advertised the study on Reddit (under the subreddit "r/PaidStudies") and Facebook (under the group "Research Participation"). Initially, we recruited 14 participants who met the inclusion criteria: individuals who (1) are over 18 years old; (2) are native English speakers; (3) have stable internet access; (4) own an Android smartphone (FoodScrap supports Android only); (5) are actively making their own food decisions (i.e., decisions on what, when and how much to eat) instead of relying on a partner or other family members; (6) are interested in collecting their food practice including food components, preparation methods, and food decisions; (7) are not practicing intermittent fasting; and (8) do not have a diagnosed eating disorder. Because we aimed to collect data at a high compliance, we excluded individuals who were practicing intermittent fasting or had a diagnosed eating disorder, who might not be able to log meals regularly.

We refined study protocol and the FoodScrap design after working with the first participant, and excluded her data for later analysis. We excluded the data of other two participants due to the data loss caused by technical issues. Therefore, we ended up analyzing the data of the remaining 11 participants (P1–11; nine females and two males). These participants lived in different regions in the US and their eating habits were influ-

ID	Age	Gender	Location	Occupation	Additional Household Members	Food Culture	Eating Goals
P1	27	F	OH	Accountant	2 Housemates	African	Eat healthier
P2	30	F	OR	Graduate student	A partner	Asian (mixed)	Increase food variety
P3	33	M	TX	Project manager	A cousin	Asian (Indian)	Boost immune system
P4	47	F	TX	Assistant writer	N/A	Asian (Chinese), American	Lose weight
P5	18	F	TX	Undergraduate student	Parents	Asian (Chinese)	Eat healthier
P6	30	F	MD	Case manager	A partner	American	Get healthier and fitter
P7	25	M	MD	Graduate student	N/A	Asian (Indian)	Eat healthier
P8	41	F	CO	Unemployed	A child	Western European	Eat Healthier and lose weight
P9	26	F	NY	Graduate student	Parents	Asian (Indian)	Eat with mindfulness and lose weight
P10	60	F	PA	Personal assistant	A partner and 2 children	American	Reduce sweets intake
P11	26	F	WA	Civil engineer	A partner	Mixed	Eat healthier

Table 5.1: Participant demographics, food culture, and eating goals.

enced by diverse food cultures. Their age ranged from 18 to 60 (M = 30, SD = 11.40). Eight participants reported prior experience using speech input on their mobile phones. Although participants were generally healthy individuals, they had specific eating goals such as eating healthier, losing weight, and reducing sweets intake. In particular, five participants reported struggling with food from time to time: P4 and P9 saw themselves as overweight, P8 and P9 thought they were sometimes emotional eaters, P10 was obsessed with sweets, and P6 tended to over exercise and had visited nutritionists regularly before the study. At the time of study, none of the participants were practicing food journaling. Detailed participants demographics was described in Table 5.1.

5.3.2 Study Procedure

The study consisted of four stages: (1) pre-study tutorial, (2) one-week data collection, (3) post-study survey, and (4) debriefing interview. At the end of the study, each participant received \$3 for capturing every main meal (i.e., breakfast, lunch, dinner) they consumed or skipped. If they captured all the three main meals they consumed or skipped every day for seven days (21 main meals), they would receive a \$7 bonus, which brought their total compensation to \$70. We applied this rewarding mechanism to encourage participants to capture as many journal entries as possible. All the compensation was provided in the form of an Amazon gift card.

5.3.2.1 Pre-Study Tutorial

We first had a one-on-one remote tutorial with each participant via a Zoom [229] video call (30 to 45 minutes). Participants were instructed to share their phone screen with me using TeamViewer QuickSupport [230], so that we could help them install FoodScrap in real-time. Before the screen sharing, we asked participants to remove any sensitive information from their home screen and to turn off all the notifications. We also shared our computer screen via Zoom, which allowed participants to see how their phone screen was displayed to us. During the tutorial, we introduced the study procedure and explained the information that participants needed to capture. We also played a video clip demonstrating how to log an entry in Main Meal Journal. In addition, we asked each participant to estimate their regular eating time for the three main meals. We then customized their reminder time according to individual's meal times right after the tutorial.

5.3.2.2 Data Collection

The next day after the tutorial, participants started using FoodScrap to capture their food practice with FoodScrap. The data collection lasted for one week, during which participants captured their meals, snacks, and skipped meals by responding to the questions asked in the three journals. All the participants met our minimal requirement for data capture: (1) capturing all three main meals (i.e., breakfast, lunch, dinner) they consumed or skipped for at least five days, and (2) capturing at least one main meal they consumed or skipped for all seven days.

5.3.2.3 Post-Study Survey

At the end of the data collection, we emailed each participant a post-study survey to measure their perceived data capture burden with FoodScrap. The survey included a set of subscales taken from the User Burden Scale (UBS) [224], which was developed to capture different types of user burden with computing systems and was later validated in many HCI studies (e.g., [108, 231]). Specifically, we employed four out of six constructs from UBS: difficulty to use, time and social burden, mental and emotional burden, and privacy burden. Refer to our Appendix B.2 for the full list of questions that we used.

5.3.2.4 Debriefing Interviews

After participants completed the survey, we conducted a semi-structured interview via Zoom with each participant. To help participants better recall their experience, we asked them to refer to their journal entries on FoodScrap by sharing their phone screen with us using TeamViewer QuickSupport. Each interview lasted 20 to 45 minutes, during which participants described their overall experience in capturing food practice with speech input. Based on participants' responses to UBS, we asked follow-up questions regarding their data capture burden.

5.3.3 Data Analysis

We analyzed participants' interaction logs on FoodScrap, journal entries, and transcriptions of debriefing interviews. We use the term *response* to refer to an answer to a single question in a journal (e.g., a journal entry contains multiple responses). Before analysis, I transcribed all the audio-recordings into text. From the interaction logs, I calculated the data capture duration of each entry as the duration between the time when the entry was started and the time when the entry was submitted, except while participants were not on the data capture interface (e.g., switching to another app).

When analyzing the responses in journal entries, we separately analyzed the responses to *meal/snack components and preparation methods* (Q4) and responses to questions on food decisions (Q5 to Q8) in Main Meal Journal and Snack Journal. We focused on the *types of details* that participants provided rather than the actual *content* of the information, because we were interested in examining the ways participants captured their food using speech rather than the types of food they chose. For the meal/snack components and preparation methods, two researchers first independently conducted Thematic Analysis [232] on the responses to identify common types of details within a subset of 247 responses (57; 23%). Through multiple sessions of discussion, we agreed on prominent detail types (See Table 5.2); then two of us independently revisited the same subset by checking which types of details they contained and reached strong agreement (Cohen's κ = .80) [1]. After resolving the discrepancies, I coded the remaining responses following the coding scheme.

For the remaining responses to the questions from Q5 to Q8, the two researchers first independently coded a subset of the 988 responses (168; 17%), and followed the same procedure as we analyzed meal/snack components and preparation methods. We categorized the responses into three groups: (1) unelaborated response, which answered the question without further explanation; (2) elaborated response, which answered the questions with explanation and examples; and (3) digression, which digressed from the original question (See Section 5.4.2 for details). This categorization follows prior works on analyzing open-ended survey responses [17, 233], which defined an elaborated response as "*additional descriptive information or explanation about a theme without introducing a new theme*" [17, 233]. We focused on examining *whether* and *how* participants elaborated their responses rather than identifying factors influencing their food decisions. During the round of revisiting the responses based on the categorization, the two researchers reached strong agreement (Cohen's κ = .70). After resolving the discrepancies, I coded the remaining responses.

We audio-recorded the debriefing interviews and transcribed them into text, and grouped the interview transcripts to answer the following questions: (1) What participants liked and disliked about using speech to capture food practice; (2) what participants' ex-

[1] Because a response can include more than one type of details or be elaborated in several ways, I calculated the value of Cohen's κ with a confusion matrix that multi-counts the responses with more than one detail or elaboration type.

perience was like in capturing their everyday food practice and how they reflected on their food decisions; (3) how participants perceived their data capture with speech input.

5.4 Results

Drawing on participants' logs, journal entries, and interview data, we report the results on: (1) descriptive statistics of journal entries, (2) descriptions on meal/snack components and preparation methods, (3) elaboration and digression in capturing food decisions, (4) benefits of speech-based food journaling, and (5) data capture burdens.

5.4.1 Descriptive Statistics of Journal Entries

We collected 275 journal entries in total, including 200 main meal entries, 47 snack entries, and 28 skipped meal entries. All but one participants captured all three meals they consumed or skipped everyday for seven days. Participants spent 148.81 seconds per session (*SD* = 97.31) capturing their main meals in Main Meal Journal, 126.41 seconds per session (*SD* = 70.71) capturing snacks in Snack Journal, and 43.71 seconds per session (*SD* = 24.28) capturing skipped meals in Skip Journal.

On average, participants generated 147.61 words (*SD* = 58.61) in Main Meal Journal, 141.61 words (*SD* = 47.49) in Snack Journal, and 48.11 words (*SD* = 26.59) in Skip Journal. In addition, we found 48 filler words (e.g., "*well*," "*you know*," "*to be honest*," "*hello*") in 45 responses, which took up 4.55% of the total responses.

5.4.2 Describing Details of Meal Components and Preparation Methods

By analyzing how participants described their *meal/snack components and preparation methods* (Q4), we identified nine different types of detail: dish names, ingredient types, individual ingredient items, spices & sauces, food portion, food characteristics, preparation types, procedural methods, and additional contexts. Table 5.2 summarizes the types of detail with descriptions, example quotes, and the number of responses in participants' journal entries.

According to our categorization, the most fine-grained way to describe a meal is explicitly listing each **individual ingredient item**, which was found in 213 (86%) responses. In the remaining responses that did not specify individual ingredient items, participants stated the **dish names** (e.g., "*salad*, " "*pizza*") or described general **ingredient types** (e.g., "*meat*," "*vegetables*," "*fruits*"). We also found that participants sometimes provided additional details regarding **spices and sauce**, **food portion**, and **food characteristics** (e.g., calorie, nutrients, taste, health values).

Most responses described general **preparation types** (e.g., homemade, from a restaurant, prepackaged, or leftover), except for a few responses that did not clearly convey this information (10 entries from 4 participants). In addition, 104 (42%) responses provided details in **procedural methods** such as cooking tools, duration, and steps.

Although question Q4 did not ask participants to provide eating contexts, we found that while describing their meals and snacks, participants naturally mentioned **additional contexts** such as people they were eating with, and how they felt about the food.

Detail Type	# of resp. (# of participants)	Description	Example quotes
Dish names	136 (11)	Commonly-used name of a dish with or without describing its components.	"*I had a Chef salad that I bought from Walmart.*" – P4
Ingredient types	13 (6)	General types of food (e.g., vegetables, fruits, meat) without specifying the ingredient items.	"*I made hard boiled dumplings meatballs, and vegetables.*" – P11
Individual ingredient items	213 (11)	Explicitly list the names of each ingredient item in the meal or snack.	"*That's an egg with no seasoning besides pepper, and then I put two slices of smoked salmon, and half an avocado.*" – P5
Spices & sauce	35 (8)	Explicitly list the spices and sauces in addition to food components in the meal or snack.	"*It had a lot of spices like powder coriander, powder cumin, spice, it has red Chilli, turmeric salt for taste.*" – P3
Food portion	30 (9)	Explicitly mention the quantity of individual food items within the meal.	"*... Two pieces of chicken, a biscuit, French fries, and a small chocolate chip cookie.*" – P1
Food characteristics	12 (3)	Explicitly describe the characteristics of the food ingredients, such as calorie, nutrients, taste, and health values.	"*... I am having a Millville Aldi's brand fiber lemon bar, and only 90 calories, which is portion controlled and I was in the mood for something a little sweet.*" – P10
Preparation types	237 (11)	Mention how the meal or snack was prepared in general, including homemade, from a restaurant, or prepackaged.	"*This is a donut I bought from Crispy Clean*" – P2
Procedural methods	104 (11)	Explicitly describe the preparation procedures, with detailed information such as cooking tools, duration, and steps.	"*... I heated it up in the microwave previously the brussel sprouts were prepared in the air fryer and the turkey was prepared in a skillet.*" – P6
Additional contexts	80 (9)	Describe the contextual information in addition to food components and preparation methods, such as how the participant felt about the food.	"*... Ever since the COVID-19 lockdown I've been trying to bake more foods. And it's been rather enjoyable.*" – P8

Table 5.2: Summary of participants' responses to meal/snack components and preparation methods (Q4) in the Main Meal Journal and Snack Journal by the type of details they provided (Note that a response can include more than one type of details).

Question	Unelaborated resp. (# of participants)	Elaborated resp. (# of participants)	Digression (# of participants)
Q5. Why did you eat at this time rather than earlier or later?	65 (11)	175 (11)	7 (4)
Q6. Why did you choose this food instead of other options?	34 (9)	209 (11)	4 (3)
Q7. When did you make the decision to eat this food?	56 (9)	182 (11)	9 (7)
Q8. Why did you eat this much food?	72 (8)	165 (11)	10 (3)
Total	227 (11)	731 (11)	30 (9)

Table 5.3: Responses to questions regarding food decisions (Q5 to Q8) in the Main Meal and Snack journals, categorized into unelaborated responses, elaborated responses, and digression.

5.4.3 Elaboration and Digression in Capturing Food Decisions

For questions Q5 to Q8 on food decisions, we grouped participants' responses into three categories: unelaborated response, elaborated response, and digression. Table 5.3 summarizes the categorization of the responses in Main Meal Journal and Snack Journal. We found that only a few responses (3%) digressed from the original question, and a majority of responses answered the questions to the point, which we considered as valid answers. Notably, 731 out of 988 responses (74%) were elaborated. In the following, we describe each category in detail.

5.4.3.1 Unelaborated Response

Unelaborated responses refer to valid answers that are high-level statements about one's food decisions without further explanation. For example, when responding to "*Why did you eat at this time rather than earlier or later?*" (Q5), unelaborated responses that were commonly logged included "*I'm hungry*"and "*It is lunch time.*" Similarly, when responding to "*Why did you choose this food instead of other options*?" (Q6), an example of unelaborated response was "*Because it is healthy*."

5.4.3.2 Elaborated Response

Elaborated responses refer to valid answers with additional information that detailed the answers. While analyzing the elaborated responses, we found that participants elaborated their responses by describing the eating moment, explaining the eating strategy, and assessing their food practice. In the following, we summarize each elaboration type (See Table 5.4 for descriptions, example quotes, and the number of responses).

Describing the eating moment: Participants expanded their responses by describing what had happened around the eating moment. The most common instances were **personal status** such as activities and feelings. In P4's statement in Table 5.4, for example, she recalled what she did before eating: "*took a long nap*," "*did a lot of work around the house*," and "*picked up my dog*," as well as how she felt: "I *was so tired*." Another common form was describing one's **food access**, especially when responding to "*Why did you choose this food instead of other options*?" Participants mentioned their food availability or constraints such as "*running out of groceries*" (P7) and "*leftover that needed to be eaten before it goes bad*" (P6). In addition, participants described how their food decisions were influenced by **social and environmental contexts**, such as people around them: "*because my mom [was] really really late, and I was actually really looking forward to this specialty from her*" (P9), and their eating environment: "*It's extraordinarily hot today in Colorado, and I have no desire to turn on the oven or stove.*" (P8).

Elaboration Type	# of resp. (# of participants)	Subtype	# of resp. (# of participants)	Description	Example quotes
Describing the eating moment	510 (11)	Personal status	271 (11)	Activities and feelings before, during, or after eating.	"*I ate it this time because I've just woke up and took a long nap. I did a lot of work around the house earlier today and I picked up my dog from the Groomer, and I was so tired.*" – P4
		Food access	188 (11)	Food availability or proximity.	"*So I'm running short on groceries, so that these are the only things that are kind of wrapped.*" – P7
		Social & environ-mental contexts	60 (11)	People around and the eating environment.	"*I have to wait until the entire family is ready to eat. So that's why we just ate at 7:40 when everyone is ready.*" – P5
Explaining the eating strategy	249 (11)	Planning ahead	108 (10)	Conscious plans regarding grocery shopping or preparation before cooking.	"*I had to do something with the chicken breast in my freezer. They needed to be defrosted. And we'll get, you know, more than one meal out of this. There will be leftover chicken sandwiches, [and] chicken with stuffing and cranberries.*" – P10
		Health beliefs	86 (10)	Belief on what one should eat to maintain a healthy diet.	"*I try to lose some weight, and they say ... I read on the internet that if you eat between the hours of 12 and 7, that you can lose some weight.*" – P4
		Habits	64 (10)	Eating routine and regular food choices that were developed over time to suit one's lifestyle.	"*This is my lunch break. Typical lunch break time at 12:30.*" – P11
Self-assessment	75 (10)	Judgment	56 (9)	Judge one's eating behavior with positive or negative comments.	"*I've been eating a lot of junk [food] so I thought I had to keep it a little [more] fresh for sustainability and health.*" – P7
		Comparison	21 (7)	Compare one's current food practice with their regular routine.	"*I would say I eat a little bit more than I normally do, but deep-fried food is something I'm into. I ate more than my normal portion but that was fine.*" – P3

Table 5.4: Summary of participants' responses to the four questions on food decisions (Q5 to Q8) in the Main Meal Journal and Snack Journal by the ways they elaborated their responses (Note that a response can be elaborated in several ways, and the elaboration types and subtypes are not mutually exclusive).

Explaining the eating strategy: Participants made food decisions based on a set of eating strategies they had specifically planned for convenience, health, or special events. Some of these eating strategies were adopted from other people or media sources, and later became participants' health belief or eating habits. The most commonly mentioned strategy is **planning ahead**. In P10's statement in Table 5.4, for example, she described how she prepared a big meal for several days. In another of P10's responses, she also explained how COVID-19 affected her eating strategies for planning ahead: "*I'm in food deliveries because of COVID. I've had to modify my diet and eat stuff like sandwiches, because my produce only lists the first week of the food order, and I'm ordering every two to three weeks for limited contact*." The second eating strategy involves participants' **health belief**. For example, in Table 5.4, P4 believed that eating between 12 to 7 p.m. can help with weight loss. In another example, P7 believed that his food was healthy because "*this is a mix of protein as well as fiber.*" In addition, participants also mentioned their **habits** including the time they usually ate, the food they regularly chose, and the amount they usually consumed.

Self-assessment: Another type of elaboration is self-assessment—expanding responses by assessing one's food decisions. A common form was to make **judgment** with positive or negative comments. For example, P10 commented on one of her snacks: "*I wanted something sweet after dinner. It's a bad habit that started [since] the last couple years*." Similarly, P7 described his lunch as "junk food." On the other hand, participants compared their current food decisions with their regular routines regarding eating time, healthiness of the food, and food amount. In Table 5.4, for example, P3 noted, "*I would say I eat a* *little bit more than I normally do*," which we categorized as **comparison**.

5.4.3.3 Digression

Occasionally, participants' responses digressed from the original questions, that is, participants provided irrelevant information or answered to another question. For example, when responding to "*Why did you choose this food instead of other options?*" (Q6), P1 responded, "*I ate this much food because this is the amount I usually eat for dinner*," which was suppose to be the answer to "*Why did you eat this much food?*" (Q8).

5.4.4 Benefits of Speech-Based Food Journaling

During the debriefing interviews, participants acknowledged that capturing their food practice using speech input was easy and fast. They also highlighted how speech input facilitated reflection on their food decisions, which we report below.

5.4.4.1 Easy and Fast Data Capture

All the participants found that speech input was easy for data capture, especially when it came to describing individual food ingredients and complicated preparation steps, as P7 remarked: "*I think filling it out via audio was much more easier than what I thought it would be. If I had to fill it out via text it would have been really difficult, because you had to mention cooking, whatever ingredients are there and everything. ... I think I would barely managed a sentence or two*." Participants' log data showed that they generally spent about two minutes completing an entry in Main Meal Journal or Snack Journal, which was perceived as time-saving by four participants: "*It's really easy and it takes less time than typing, I think.*" (P11).

5.4.4.2 Speech Journaling as a Reflection Tool

Before the study, participants had rarely consciously thought about when and why to choose what to eat. Therefore, responding to the journal questions helped participants become better aware of the relationships between their physiological feelings and their eating behavior. For example, participants sometimes were surprised to find out how their food decisions differed from what they had believed: "*I was surprised this week at how many times I was really just eating because I was hungry. I thought I was a much more emotional eater.*" (P8). In particular, P2 emphasized that speaking out her food decisions made her eating patterns more noticeable: "*When I answer that question 'why did you eat at this time' I learned how sporadic our eating is like. [...] I was saying those things, which kind of made it more obvious.*" Interestingly, P-10 said "*hello*" and "*good morning*" in many of her journal entries like she was interacting with a real person. She explained, "*I would say hello, or good morning, because I'm extremely outgoing and I'm very verbal. [...] Even though I was talking into an electronic [phone], I feel like interacting with people, so it made me want to talk more. I feel more accountable, you know, to explain my food [decisions], to really think about it, like why am I eat this now*."

While capturing food decisions in the process of eating, participants started thinking about their eating behaviors in a more mindful way and even tried to regulate their eating intention. For example, responding to "*Why did you eat this much*?" nudged P11 to stop and to ask herself: "*Do I really want to eat the whole bag of chips*?" Similarly, P-9 remarked: "*I mostly just use it [FoodScrap] as a tool for my self-reflection, I guess I overthink things all the time, and I always reflect on what I said. So sometimes I thought maybe I should stop [eating]*."

In addition, participants had distinctive preferences on whether to listen to their audio recordings. Seven participants never played back their recordings because "*I don't like my voice*." (P6). P9 also added that "*because I don't listen, so I can speak whatever I thought of*." On the contrary, four participants would play back their recordings to check the audio quality and to reflect on past eating episodes: "*I did this for checking the quality of the audio. Also sometimes I'm curious how much my food decisions were influenced by others versus myself*" (P5). In P11's case, although she listened to the recordings without specific purposes, she valued the convenience of revisiting past food decisions with no need to focus on her phone screen: "*I wasn't looking for something specific. I think it was just easy to listen and you don't need to keep your eyes on the screen, and there will be moments like oh, that's what I was thinking back then*."

5.4.5 Data Capture Burden

The average User Burden Scale (UBS) score across the four metrics—-difficulty to use, mental & emotional burden, time& social burden, and privacy burden—were relatively low (between 0 to 1)[2], indicating that the speech-based data capture burden was low. However, during the debriefing interviews, participants reported concerns around re-

[2]Scale ranges from 0: "No burden at all" or "Never (happened a burdensome situation)" to 4: "Extremely burdensome" or "All of the time (it was burdensome)"

recording effort, mental load, social constraints, and privacy. In the following, we share examples regarding these types of data capture burden.

5.4.5.1 Re-Recording Effort

Four participants reported that sometimes they had to re-record their responses if they lost the train of thought in the process of recording, which took more time than expected: "*I'd be like talking about what I ate, ... You know, I would start talking about something else, and then I'd be like, Oh no, this is not responding to the full question. So then I'll delete it, and then redo it. So sometimes it took like a little bit more [time].*" (P2). Although FoodScrap provides a "pause" option that allowed participants to manipulate their recording progress, they seldom used this option; instead, participants preferred deleting the entire audio to start over: "*When I was disturbed, I wasn't able to complete my sentence. Pausing doesn't help, so I deleted the recording altogether.*" (P4).

5.4.5.2 Mental Load

Participants reported that journaling with speech input sometimes required extra attention and concentration, especially in two cases: when they ate mindlessly without clear answers to the questions or when they had a lot to say about their food decisions. Four participants mentioned that they felt difficulty in responding to the questions on food decisions because of mindless eating: "*Most of the time I found myself eating, and I couldn't really tell why, why I ate at this time, or why I chose this food. I felt it's hard to give an answer, it might be just an intuition, or like a habit, but I can't explain why.*" (P1).

On the other hand, P5 and P7 often needed to think through and organize what they wanted to speak before recording their responses. To make sure that their responses were clear and concise, it usually took extra mental load: "*Because I don't want to record [an] audio for a minute or two, where I'm fumbling through my sentences. So I needed to gather my thoughts regarding what I need to say quickly. So initially, it was a little jam regarding what I wanted to say.*" (P7).

5.4.5.3 Social Constraints

Participants reported being constrained by social contexts while using speech input, especially when other people were around. Three participants expressed that they felt embarrassed talking to their phones in a public space: "*I also need to think about when I'm going to record, because sometimes there are others present. It's weird picking up my phone and talking to it.*" (P11). Other two participants expressed concerns about including surrounding noise in their recordings: "*One time I had to go in the bathroom, because my daughter was having a play date and they were just kind of being noisy, so I had to bring my phone in the bathroom and make the recording.*" (P8).

5.4.5.4 Privacy Concerns

Three participants considered food practice to be private, and were concerned about their food decisions being judged by others. Therefore, they raised concerns on disclosing their food practice through speech input because "*voice is more identifiable than text*" (P5). For example, P9 mentioned that she was very self-conscious preventing people around from hearing what she spoke to FoodScrap: "*I know the study doesn't judge my habits, I was concerned about what others around me might judge how I was eating. So I would have to make sure that I was in a relatively private place, so that I could speak clearly and wouldn't be overheard on*."

5.5 Implications

In this study, we showed that speech-based input is promising in lowering the data capture burden while promoting situated reflection. However, it is important to consider how to process and present the large amount of speech input so that they can be useful for self-trackers, healthcare providers, and researchers. Furthermore, more work needs to be done to address the constraints that come with speech-based input to support data capture in different social contexts.

5.5.1 Collecting Rich Details Through Fast and Expressive Data Capture

Our participants provided rich details in their food components and preparation methods, which could be laborious to capture via touch-based typing, as P3 explained while showing one of his journal entries: "*This is a 45.7 second recording that I did. Now imagine, if I need to type, that would be too much writing. I'll probably miss some data or try to cut corners with it*." We note that many of the details—such as condiments and preparation procedure—are critical information for assessing meal healthiness [234–236] but are difficult to capture through retrospective surveys or even automated food recognition technologies (e.g., photo, barcode) [237]. In dietary assessment, for example, dietitians and nutritionists often employ dietary history method [234] and food frequency questionnaire (FFQ) [235], which ask for more than 90 items about one's food intake, covering details such as "*seasonings and flavorings*" and "*cooking methods*," but do not always produce accurate results due to the time lag [234]. Our study suggests that speech-based data collection can capture more details *in-situ* with lower data capture burden, which may improve the data accuracy [89]. However, to fully leverage the large amount of speech data, we need to consider how to efficiently process and present the data for healthcare providers' use. With the advances in natural language processing (NLP), we can extract food-related information (e.g., food group, portion size, ingredients) from the transcribed text [100] and support customized information sorting & filtering based on providers' needs [218].

When answering questions on food decisions, participants often elaborated their responses, which resonates with prior research suggesting that people tend to be expressive when they are speaking [14]. These elaborated responses are usually ephemeral and momentary contexts—personal status, food access, and social and environmental contexts around the time of eating—that are valuable information for dietitians and food science researchers, but can be hard to capture through retrospective recall. For example, understanding how patient's living environment and social life shape their food decisions helps dietitians deliver more personalized care: dietitians may help patients restructure their eating environment instead of simply prescribing what to eat [238], or use food journal as an intervention to encourage mindful eating [218, 239]. While current practice of understanding food decisions often relies on verbal communication during clinical consultation [240], FoodScrap enabled participants to capture food decisions that are tied to every meal or snack, providing opportunities to capture rich details that might otherwise be overlooked.

Our findings demonstrated the potential of speech input to capture detailed food information and elaborated food decisions that are typically hard to capture through other approaches (e.g., typing, automated means, interviews). In this regard, speech input holds promises in other self-tracking contexts beyond food journaling (e.g., capturing perceived workout intensity and feelings in exercise tracking), where individuals and researchers can identify nuanced but important insights from one's daily activities [241, 242].

5.5.2 Fostering *Reflection-in-Action* Through Guided Prompts

Self-tracking technologies support reflection in various ways [243]: providing real-time feedback (e.g., [26]) or augmenting manual data capture(e.g., [11, 90, 218]) can support reflection-in-action; and providing aggregated feedback of past behaviors (e.g., [45, 107, 244, 245]) can support reflection-on-action. In our study, FoodScrap mainly facilitated reflection-in-action at the time of data capture. Among participants' elaborated responses, we found several instances involving self-assessment with *judgements* or *comparison*, which were indicators of reflection-in-action [246]. Those reflective thoughts were likely resulted from the guided prompts in FoodScrap, which questioned participants to think about their food decisions in specific aspects such as when and how much to eat, and why they choose the food. Furthermore, we suspect that speech input might have nudged reflective thinking by supporting free-form expressions, as P-10 remarked that thinking aloud was like "*interacting with people*," which made her feel "*more account-*

able" to explain her food decisions. This finding corroborates a previous study in which researchers found that video recording of eating experience with narration could promote self-reflection through contextualizing one's eating experiences [90]. Such free-form expression is important for people who struggle with food (e.g., eating disorder patients) to raise situated awareness and to build positive self-image [218].

As reflection-in-action happens during the moment of data capture, which is close to the time of eating, we see opportunities for encouraging mindful eating during these "critical reflection moments" [114, 218, 239]. For example, asking "*why do you want to eat now?*" may prompt people to think twice about their decisions and to be more mindful about whether their cravings are caused by hunger or boredom [218]. To understand how different modalities of data capture (e.g., speech recording, video recording) support reflection-in-action, future work remains to compare these modalities with traditional text input or other structured entry forms.

5.5.3 Enabling *Reflection-on-Action* Through Feedback on Past Data

To fully support a reflective food journaling experience, it is important to enable reflection-on-action through delivering aggregated or summary feedback of past behaviors, so that individuals can stay engaged by reflecting on the patterns of their food practices [11, 45]. The focus of the FoodScrap study was on the data capture aspect, so we did not provide any feedback beyond the capability of replaying the audio. While the unstructured nature of speech input adds complexity to data processing and analyzing, we see opportunities for presenting the rich information in both visual and auditory forms.

We can summarize individuals' responses corresponding to each guided prompt as feedback. For example, the most commonly mentioned reasons that affect one's eating time, food choice, and amount of food consumption can be shown in text summarization using keyword extraction and term-frequency analysis. To further support individuals exploring their data, the keywords extracted can be visualized in a word cloud [247].

Along with the responses to questions on food decisions, participants gathered information beyond what we asked: they described how they felt about the food, how they planned for other meals, and how they assessed their food practice, etc. In particular, we found many instances related to participants' emotional feelings. For example, one of P9's responses—"*I ate this much food because I felt depressed and didn't know what to do with myself. Honestly, so I just finished the whole part in one session.*"—indicates that the feeling of depression could have caused her to eat more food than she needed. In such cases, we can use sentiment analysis to identify emotion-related information, and help individuals draw insights on how their emotion (e.g., positive and negative) may be related to their food decisions.

Four (36%) participants in our study replayed their audio recordings and valued their recordings as resources for revisiting past eating episodes. This finding implies the potential of auditory feedback to support reminiscence and reflection, which can be important for those who track their mood, stress, and mindful thoughts [27]. We suspect that when people audio record short and structured data consisting of numbers or simple phrases, text summary or chart is a better form than auditory feedback for reviewing purposes. On the other hand, when people capture long and complicated information, retaining the original audio recording could be valuable [248], as it might contain unique contextual information that text transcription cannot provide, such as pitch, tone, and volume of the voice as well as background sounds. To enable more efficient audio searching, we can provide text summary (e.g., extracted key words) or visual feedback (e.g., photos) along with the original audio recording.

5.5.4 Supporting Data Capture in Varying Contexts Leveraging Multimodal Input

While the UBS score indicated that the overall data capture burden with FoodScrap was relatively low and all the participants acknowledged that speech input was easy and fast, we noticed that leveraging speech for capturing complex and long information is not always desirable. As participants expressed concerns around social constraints and privacy, speech-based data capture seemed to work better in a private setting rather than a public setting.

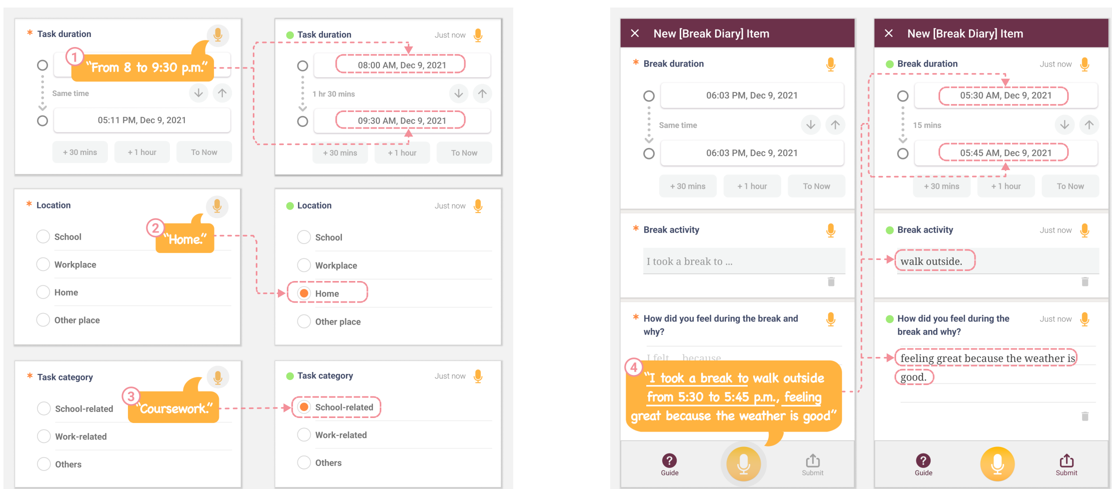
To support food tracking in varying contexts, we can leverage multimodal input combining speech, text, and photo across multiple devices (e.g., smartphones, smart speakers, wearable devices, wireless earphones) so that people can choose *when* to use *which* input modality. For example, in a privacy-sensitive situation (e.g., crowded place, office setting), people may choose text input on a smartphone; at home where the smartphone is not close by, people can use speech input on a smart speaker or wearable devices with the hands-free interaction [91]. In another case when people do not have enough time to capture all the information at once, they can take a food photo first, and add more details afterward using speech or text.

In addition, participants reported occasions where they eventually spent extra time on re-recording their responses when being disrupted or losing the train of thought. One potential solution is to provide real-time transcription, which may help individuals keep their train of thought and reduce mental load. If people are not satisfied with their responses, they can edit the data by typing instead of re-recording the entire response.

5.6 Chapter 5 Summary

Chapter 5 describes a week-long data collection study with FoodScrap, a speech-based food journaling app that we created to capture food components, preparation methods, and food decisions. Throughout the study, 11 participants collected rich data, including detailed information about their food intake and elaborated statements of their food decisions. I distilled the ways that participants used speech input to describe their food practice, and summarized speech input's benefits and drawbacks regarding data capture burden. In particular, I highlighted speech input's fast and expressive data capture in collecting flexible and nuanced details and its potential for fostering reflection-in-action. I also discussed opportunities for leveraging speech input to further support reflection-on-action, and designing multimodal input systems to facilitate data capture in varying contexts. In summary, this work contributes to an empirical understanding on how speech input supports capturing unstructured self-tracking data and informs the design of multimodal self-tracking tools to capture rich data in a low-burden and reflective manner.

Chapter 6: NoteWordy: Investigating Touch and Speech Input on Smartphones for Personal Data Capture



(a) Local speech (LS) input (b) Global speech (GS) input

Figure 6.1: NoteWordy integrates touch and speech input to support people to capture different types of data. With touch input, people can pick time points, select multiple choices, and type text. With speech input, they can capture a single data field by pressing on the ***local speech*** (LS) button placed on that field 1 2 3, or multiple data fields together by pressing on the ***global speech*** (GS) button at the bottom center 4 (the keywords that helped the system segment and extract the information are underscored). Please refer to our supplementary video for interaction details.

In Chapter 5, I examined how speech input on mobile phones supports people to capture long and unstructured data about their everyday food practices. However, it still remains unclear how we can incorporate speech input to work with other input modalities on the same device. In this chapter, I designed and developed NoteWordy, a multimodal self-tracking app equipped with touch and speech input to capture different types of structured and unstructured data. By deploying NoteWordy in the context of productivity tracking, I aim to answer **RQ3**: How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes? and **RQ4**: How does the input modality affect the data richness of unstructured (i.e., free-form text) input?

6.1 Introduction

Self-tracking in real-world settings often involves capturing multiple data in different types [3,117]. A typical example is tracking daily activities by capturing time, location, activity type, and other contexts [10, 11, 21]. However, the majority of self-tracking tools employ touch input interfaces requiring people to perform a series of manual selections and typing, which imposes a heavy data capture burden.

This work examines the underexplored potential for speech input to facilitate multiple data capture: instead of entering each data item one by one, people can include multiple data in one sentence through speech input. For example, the sentence "*I was at home at 9 a.m., having some coffee and feeling refreshed*" captures one's location (home), time (9 a.m.), activity type (having some coffee), and feelings (refreshed). In the meantime, the flexibility of natural language allows people to phrase the same information differently. For example, they can describe time-related information with standard timestamps, relative time points, or special holidays and events [16]. Furthermore, in capturing unstructured data in free-form text, the expressive nature of speech input can enhance the data richness by encouraging people to provide additional details [249]. However, prior research has not yet systematically examined whether and how speech input affects the data richness compared with manual typing.

In this light, my colleagues and I designed and developed NoteWordy, a multimodal mobile app integrating touch and speech input to capture different types of data (Figure 6.1). NoteWordy allows people to manually capture their data with touch input and offers two speech input options: ***local speech*** (LS) input for entering one data field at a time by pressing on the LS button placed on that field (Figure 6.1a); and ***global speech*** (GS) input for entering multiple data fields at once by pressing on the GS button at the bottom center (Figure 6.1b). With touch, LS, and GS input, people can capture multiple data fields individually or together with their preferred input modality.

We situated this work in the context of productivity tracking, because productivity can be characterized by multiple dimensions (e.g., task duration, work output, and mental status) in different data types [131], allowing us to examine the research questions. We targeted working graduate students—individuals who attend graduate school and work off-campus concurrently, because (1) unlike undergraduate students with structured course schedules and GPA-oriented goals [250], graduate students tend to have more flexible schedule but may experience more stress due to career transition, financial burdens, or family obligations [251]; and (2) for those who are also employed for another job, time management can be even more challenging and complicated. Therefore, working graduate students are likely to be interested in collect productivity-related data for better self-understanding and time management [131]. The collected data can also help researchers and educators develop tailored coping strategies and productivity tools for this particular group [252–255].

We created two diaries: "Productivity Diary" and "Break Diary" in NoteWordy, and deployed the app to 17 working graduate students for two weeks. During the study, participants could choose between or combine touch and speech input to capture data about their tasks and breaks. With both quantitative and qualitative approaches, we summarized participants' input patterns with touch and speech input, and investigated how speech input affected the diary completion time and data richness in unstructured input. In the following, I describe the design and implementation of NoteWordy and findings from the two-week data collection study. In addition, I discuss opportunities to improve the data capture experience combining touch and speech input.

6.2 NoteWordy

NoteWordy was built upon the client app of OmniTrack for Research [225], which already supports capturing different types of data with touch input. Our design and implementation of NoteWordy thus focused on incorporating speech input to capture individual and multiple data fields. In the following, I first describe our design rationales, and then present NoteWordy's speech interface along with implementation details.

6.2.1 Design Rationale

6.2.1.1 DR1: Provide Both Touch & Speech Input Capabilities

According to prior research, people have individual preferences for the input modality [91] while their choices also being affected by external factors such as social environments [91, 249]. Instead of designating speech input for specific data fields, we aim to understand how touch and speech work together to support capturing different types of data (RQ3). Therefore, we made both input modalities available for each data field to let people choose between or combine the two input modalities in the way they like.

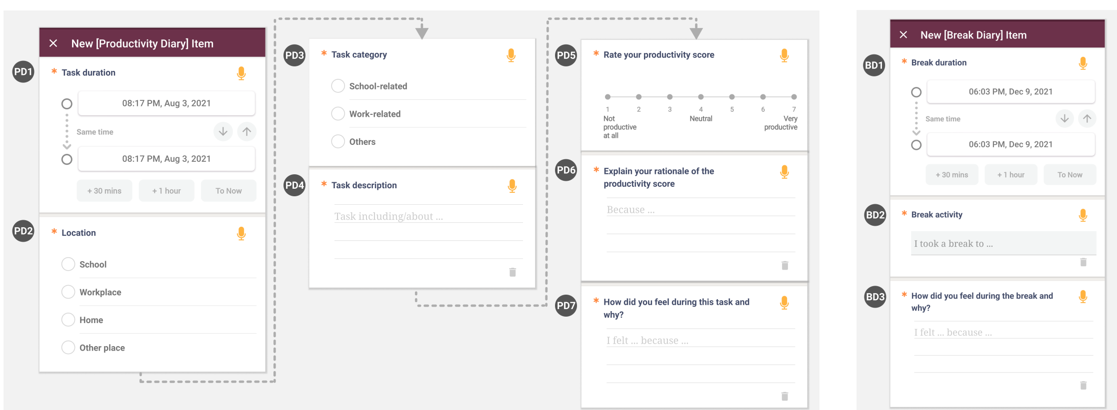
6.2.1.2 DR2: Enable Flexible Data Capture With Natural Language

As a natural input, speech allows people to capture the same data with different expressions [16, 256]. For example, they can select an item in multiple choice questions by saying its synonyms and capture time points in standard or (e.g., "*8 in the morning*") relative (e.g., "*two hours ago*") forms. Furthermore, People can capture multiple data fields without following a particular order in one utterance. In the example illustrated in Figure 6.1a, task duration, location, and task category can also be captured in one sentence, such as "*I did coursework from 8 to 9:30 p.m. at home*", instead of individually saying "*from 8 to 9:30 p.m.*," "*home*," and "*coursework*." Therefore, in addition to supporting people to individually capture each data field with speech input, we also aim to support them to capture multiple data fields together using a variety of expressions.

6.2.1.3 DR3: Design for Clear Speech Guidance

One main challenge that people often face with speech interface is the discoverability (i.e., the ability to discover the correct speech commands to interact with the system) [257]. Without clear guidance, people are unsure about how to phrase an utterance [91, 258]. As a result, they may end up abandoning speech input and turning to other modalities (e.g., touch) [91]. To help people understand the capabilities and limitations of NoteWordy, we aimed to guide them through the input process by providing utterance examples when initiating the speech input. We realized that the current speech recognition and data processing techniques are not perfect due to the complexity of natural language [256] and various external noises (e.g., background sounds, microphone quality) [259]. Thus, when a speech recognition error occurs, it is important to inform people of what caused the error and what alternatives they can try.

6.2.2 Data Capture With NoteWordy



(a) Productivity Diary (b) Break Diary

Figure 6.2: The questions asked in Productivity Diary (a): task duration (PD1), location (PD2), task category (PD3) and description (PD4), productivity score (PD5) and rationale (PD6), and feelings during the task (PD7); and in Break Diary (b): break duration (BD1), break activity (BD2), and feelings during the break (BD3).

6.2.2.1 Diary Design: Productivity Diary & Break Diary

Drawing from prior research on productivity data collection, we focused on three aspects that play important parts in one's daily productivity: tasks [131], breaks [58], and mental status (e.g., feelings) [190]. We created two diaries in NoteWordy: Productivity Diary, which captures task-related activities and feelings, and Break Diary, which captures break-related activities and feelings. In Productivity Diary (Figure 6.2a), we first examine how people spend their time by asking their task duration (PD1), task location (PD2), task category (PD3), and detailed task description (PD4). We also asked people to rate their productivity score in a Likert scale from one to seven (PD5), a metric that has been frequently used in previous work [23, 260]. To further understand how people evaluate their productivity, we added a question asking them to explain the rationale of the productivity score (PD6) in free-form text. Lastly, we asked people how they felt during the task and why (PD7) to capture their mental status. In Break Diary, we shortened the questions to focus on people's break duration (BD1), break activity (BD2), and how they felt during the break and why (BD3) (Figure 6.2b). These questions cover four types of data: timespan (PD1, BD1), multiple choice (PD2, PD3), Likert scale (PD5), and free-form text (PD4, PD6, PD7, BD2, BD3), allowing us to investigate how people use touch and speech input to capture different types of data.

6.2.2.2 Local Speech (LS) Input

To provide both touch and speech input capabilities (DR1), we placed a ***local speech*** (LS) button on each data field (Figure 6.1a). With the "push-to-talk" operation, the system records the speech input while people are pressing on the LS button of the specific data field that they intend to capture. The system handles natural language input (DR2), allowing people to record the data in the ways that they are familiar with. In PD1, for

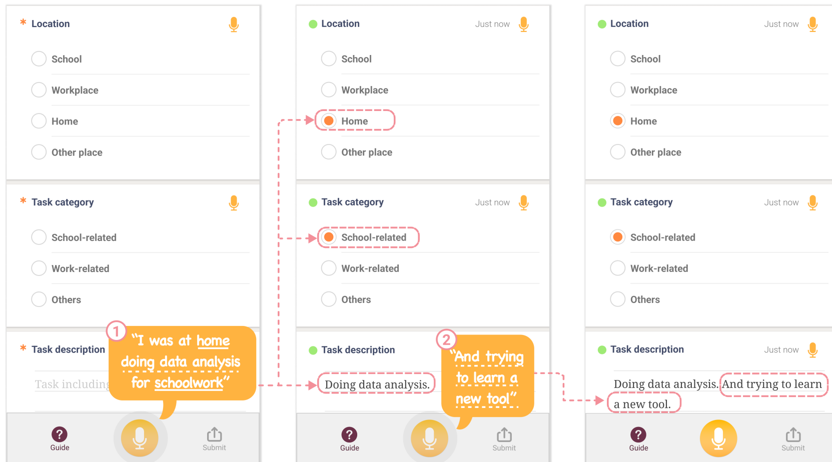


Figure 6.3: An example of how GS handles uncategorized text (Productivity Diary): if no data field in the current view was filled or the last filled data field is not a text field, the uncategorized text will be inserted into to the next text field coming along ①; if the last filled data field is a text field, the uncategorized text will be appended to that text field, allowing people to incrementally add information to the same text field with GS ②. (The keywords that helped the system extract the information are underscored with solid lines and the unrecognized text is underscored with dotted lines).

example, people can provide the start and end time or mention the duration with only one of the two points. In PD2 and PD3, they can select an item by describing the item name or using similar phrases (e.g., say "*company*" to refer to "*workplace*" in location). To capture productivity score in PD5, people can say a number from one to seven or a label from "*not productive at all*" to "*very productive*" (e.g., say "*productive*" to refer to "*6*"). When people press on the LS button on any of the text fields (PD4, PD6, PD7), all the transcribed text from their speech input will be entered into that field. They can also append more text to that field by pressing on the LS button and speaking again.

6.2.2.3 Global Speech (GS) Input

We provide a ***global speech*** (GS) button  that is unattached to any data fields (Figure 6.1b) so that people can capture multiple data fields at once (DR2). GS also adopts the "push-to-talk" operation that records speech input while people are pressing on the GS button (e.g., press on the GS button to capture task category and description by saying "*I had a work meeting about the mockup design with the engineering team*"). People are asked to include certain keywords in their utterances to help the system extract the key information. The recommended keywords for each text field are displayed in gray text as a hint (e.g., "*Tasks including/about* ..." under PD4 (Figure 6.2)). To improve the accuracy of the extracted data, we appended several synonyms of each keyword in the systems' vocabulary (e.g., "*because* ..." can be replaced with "*due to* ...").

NoteWordy handles the uncategorized text segment from the GS input (i.e., text segment that either belongs to any structured data fields or includes any keywords of existing text fields) in two ways: (1) in the current view, if no data field was filled or the last filled data field is not a text field, the uncategorized segment will be inserted into the next text field coming along (Figure 6.3 ①); (2) if the last filled data field in the current view is a text field, the uncategorized segment will be appended to that text field, allowing people to incrementally add information to the same text field with GS (Figure 6.3 ②). The rationale was that people are likely to complete the diary entries by following the order of the questions and may need to enter data into the same text field multiple times. In case that the uncategorized segment is placed to the wrong field, we provided a clear button at the right bottom of each text field, allowing people to easily delete the text and start over.

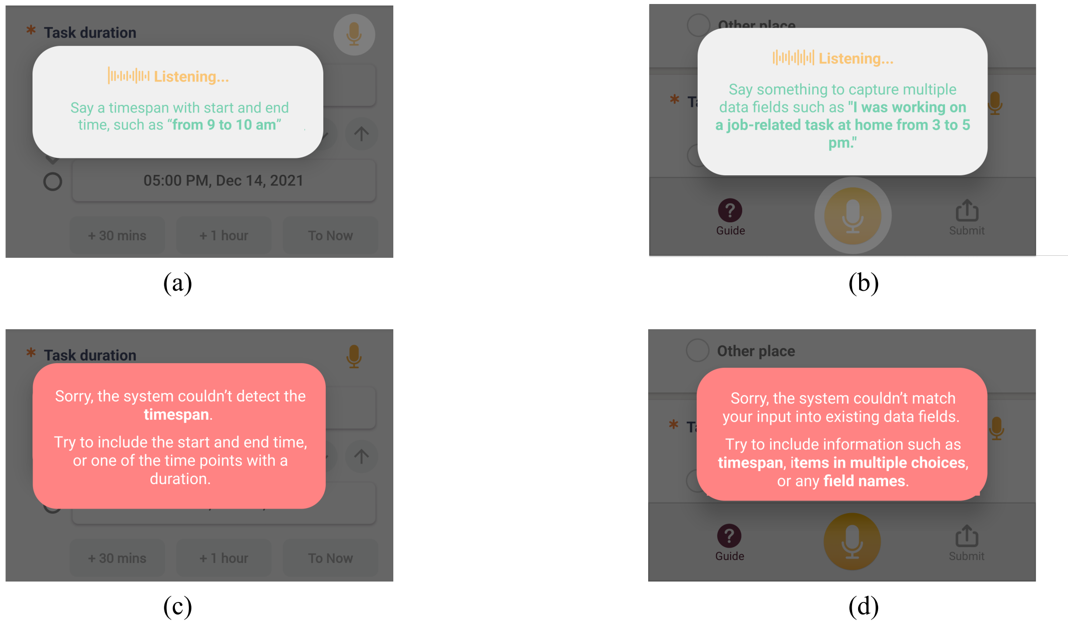


Figure 6.4: The speech input dialog that displays contextual messages to guide people through the recognition process: (a) When people press on the LS button placed on the Task duration field; (b) when people press on the GS button with Task duration, Location, and Task category fields visible in the current view; (c) when the system fails to recognize the input for Task duration; (d) when the system fails to recognize the input from GS.

6.2.2.4 Contextual Guide & Error Feedback

When people press on the GS button or any of the LS buttons, a speech input dialog pops up to guide people through the recognition process while dimming the screen behind (see Figure 6.4). Before people start talking, the dialog displays a contextual message explaining what they can say with an utterance example, which is based on the data field that the button is placed on (e.g., showing "*from 9 to 10 am*" for Task duration field (Figure 6.4a)). When the GS button is pressed, the dialog displays an utterance example based on the data fields that are visible in the current view (e.g., showing "*I was working on a job-related task at home from 3 to 5 p.m.*" when Task duration, Location, and Task category fields are all on the screen (Figure 6.4b)). While people are talking to NoteWordy, the dialog displays live transcripts of people's speech input so that they are aware of how their utterance is being recognized. This also prevents people from releasing their finger before the system completes recognition. If NoteWordy fails to recognize the speech input, an error message pops up to inform people of what might be wrong and suggests alternative utterances that they can try (Figure 6.4c, 6.4d).

6.2.3 Implementation

Extending OmniTrack, NoteWordy is written in Kotlin [261] on Android platform. We used Microsoft Cognitive Services [262] as a speech-to-text recognizer instead of Android's built-in speech recognizer because (1) Microsoft's service allows developers to customize timeout for continuous recording, so that we could avoid potential problems caused by speech input being cut off when people pause in the middle of recording; and (2) the service provides automatic punctuation, which helped with easier text segmentation. We used SmileNLP [263], a machine learning engine to further segment the utterances and to handle different forms of the same word (e.g., "*feeling*" and "*felt*" are different forms of "*feel*"). We also incorporated a time parser called Natty [264] to process different time expressions. To improve the recognition accuracy, we appended a set of keywords related to the study context (e.g., "*coursework*" and "*schoolwork*" are synonyms for "*school-related*") to the speech recognizer's vocabulary. The pipeline that processes the speech input from GS is illustrated in Figure 6.5.

The data collected by NoteWordy are securely stored on a virtual machine hosted on the university's server. People can access their data in the app by revisiting the raw

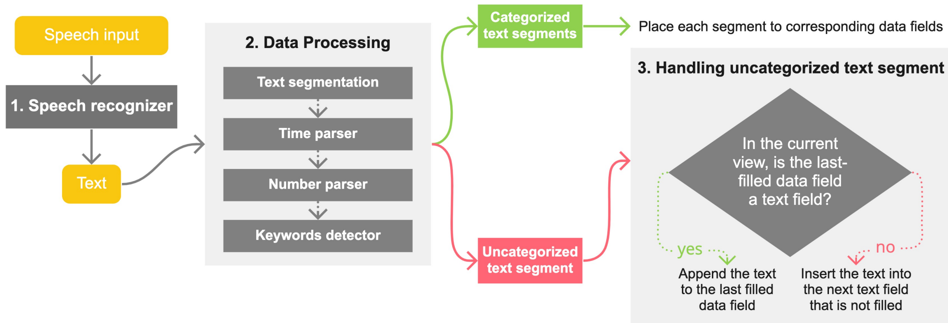


Figure 6.5: The pipeline that processes speech input from GS: 1. Transcribing speech input into text; 2. Extracting the structured data from the text and categorizing other text segments based on the keywords; 3. Handling uncategorized text segment.

entries or aggregated visualizations (e.g., the number of daily entries, productivity score across time). The details of the visualization design are described in [21].

6.3 Methods

We deployed NoteWordy to 17 participants for two weeks, followed by debriefing interviews. All the interactions we had with participants were remote on Zoom [229]. The study was approved by the university's Institutional Review Board (IRB # 1132164-23) and conducted from August to September in 2021.

6.3.1 Participants

We advertised the study through the university mailing-list, Reddit (under the sub-reddit "r/GraduateSchool" and "r/MBA"), and Facebook (under the group of "Graduate School"). We also approached several Reddit users who posted discussions about time management as working graduate students, and asked if they would like to participate in the study. Our inclusion criteria were adults who (1) are fluent in English; (2) possess an Android mobile phone with an OS version 4.4 or above; (3) are enrolled in a graduate

ID	Age	Gender	Student type	Major	Off-campus occupation	Employee type	Work mode	Experience with speech interface
P1	27	M	Full-time master	MBA	UX Designer	Full-time	Remote	Neutral
P2	30	M	Part-time master	Data Science	IT Administrator	Full-time	Remote	Positive
P3	30	F	Part-time master	Data Science	Data Analyst	Full-time	Remote	Positive
P4	24	M	Part-time master	HCI	Data Engineer	Full-time	Remote	Positive
P5	25	F	Full-time Ph.D.	Computer Science	Art Designer	Freelancer	Remote	Neutral
P6	35	M	Part-time master	Statistics	Research Analyst	Full-time	Hybrid	Positive
P7	26	M	Part-time master	MBA	Photographer	Part-time	Hybrid	Positive
P8	30	F	Full-time master	Library Science	Music tutor	Part-time	Hybrid	Neutral
P9	24	F	Full-time master	Medicine	Behavior technician	Part-time	In-person	Positive
P10	22	M	Full-time master	Computer Science	Researcher	Part-time	In-person	Positive
P11	35	M	Full-time master	Theoretical Physics	Database Operator	Part-time	Hybrid	Positive
P12	37	F	Part-time master	Library Science	Research Analyst	Full-time	Remote	Neutral
P13	25	M	Full-time master	HCI	Newsletter Coordinator	Part-time	Remote	Neutral
P14	26	F	Full-time master	Classics	ESL Instructor	Part-time	Hybrid	Positive
P15	28	F	Part-time Ph.D.	Aerospace Engineering	Aerospace Engineer	Full-time	Remote	Positive
P16	29	M	Full-time Ph.D.	Computer Science	Researcher	Part-time	In-person	Neutral
P17	37	F	Full-time Ph.D.	Social Science	Researcher	Full-time	Hybrid	Neutral

Table 6.1: Participants' demographic, student and employment types, work mode, and experience with speech interface.

program at a university (master's or Ph.D. level); (4) are employed full-time or part-time outside the university, working at least 20 hours per week in addition to schoolwork; (5) are curious about how they spend time between school and work; (6) have no visual, motor, or speech impairments; (7) have experience using speech interaction and are willing to use it daily; (8) have stable access to the Internet; and (9) have a computer with a webcam, microphone, and speaker so that they can communicate with the researchers via video chat. By looking for graduate students who worked off-campus, we excluded those who were graduate research/teaching assistants, because their work is often a part of their school credits or overlaps with their school tasks, thus the boundary between their work-related and school-related tasks may not be clear. We excluded people with disabilities (i.e., visual, motor, or speech impairment) because the design of NoteWordy did not specifically consider the special needs of these groups.

We initially identified and contacted 66 qualified individuals, and 27 of them replied to our email. During the pre-scheduled tutorial, 23 participants showed up on Zoom and 20 of them completed the tutorial; the other three could not complete the tutorial due to technical issues. At the end of the study, 17 participants completed the data collection, and the other three dropped out because of busy schedules and being unresponsive during the study. The 17 participants' (P1–17; 8 female and 9 male) age ranged from 22 to 37 (*Median* = 28, *SD* = 4.7) and lived in different regions in the US (see Table 6.1). Thirteen of the participants were master's students (6 part-time) and four were Ph.D. students (1 part-time). Our participants majored in different fields of study (e.g., MBA, computer science, medicine, classics) and had different jobs (e.g., UX designer, data engineer, photographer, researcher). All the participants had used speech interface (e.g., voice assistant on their phones, smart speaker) before, and 11 of them were positive about the experience. Three participants (P2, P3, P17) reported themselves as working student-parents, and two participants (P2 and P6) reported that they had been diagnosed with attention-deficit hyperactivity disorder (ADHD).

6.3.2 Study Procedure

The study consisted of three stages: (1) tutorial, (2) two-week data collection, and (3) debriefing interviews (optional). After completing the data collection, each participant received a $30 Amazon gift card as compensation. Those who opted in to the debriefing interviews received an additional $10 gift card.

6.3.2.1 Tutorial

At first, we had a one-on-one remote tutorial with each participant (45 minutes). We asked participants to share their phone screen with us via TeamViewer QuickSupport [230] so that we could instruct them to install NoteWordy and watch them interact with the app in real-time. We then shared our computer screen via Zoom so that participants can see how their phone screen was being displayed to us. Prior to screen sharing, we asked participants to remove any sensitive information from their home screen and turn off incoming notifications to mitigate the risks of accidental privacy disclosures.

During the tutorial, we went through the study procedure and described the types of data that participants needed to collect. After demonstrating how to enter each data field with LS and GS, we led a short practice session with each participant. First, we asked the participant to enter the data fields, individually with LS and together with GS, by following the example utterances we prepared. Next, the participant was allowed to freely explore the touch and speech input to get familiar with the interface for two to three minutes. The participant then needed to think about a recent task and a break, and complete one entry respectively in Productivity Diary and Break Diary. Lastly, we explained that

NoteWordy's speech recognition was not perfect (e.g., missing keywords) and situations where recognition issues might happen (e.g., talking too far away from the microphone).

6.3.2.2 Data Collection

The data collection started the next day after the tutorial and lasted for two weeks, during which participants used NoteWordy to capture their tasks and breaks. To ensure that participants capture their tasks across different times during the day, we segmented the daytime into four windows: 9 to 12 p.m., 12 to 3 p.m., 3 to 6 p.m., and 6 to 9 p.m. As a minimal requirement of data capture in Productivity Diary, each participant needed to capture one task in three of the above time windows per day (e.g., one task from 9 to 11 a.m., one task from 4 to 6 p.m., and another task from 6 to 7 p.m.). In Break Diary, participants needed to capture at least one entry per day. Our study focused on capturing "intentional breaks" that participants took to refresh and relax instead of "unintentional breaks," such as being distracted by social media or going to the bathroom. During the 14 days, each participant was allowed to skip their daily entries for 2 days. At the end of the data collection period, 17 of 20 participants who participated in the tutorial met the minimal requirement of data capture.

6.3.2.3 Debriefing Interviews

Upon the completion of data collection, we contacted each participant to ask if they were interested in attending a remote interview for 30 to 45 minutes. To help participants recall their study experience, we asked them to open their diary entries on NoteWordy and share their phone screen with us using TeamViewer QuickSupport. All the participants opted in to do the interview, during which we asked questions about how they chose between or combined touch and speech input in different scenarios, their preferences for LS, GS, and touch input, and the challenges they faced in completing their diary entries.

6.3.3 Data Analysis

The study generated a mix of quantitative and qualitative data, including participants' interaction logs with NoteWordy, diary entries, and subjective feedback from the interviews. Here, I describe how we analyzed these data to answer the research questions.

6.3.3.1 Log Data Analysis

We first summarized the descriptive results of participants' diary entries and the input modalities that they used to capture each data field. We then looked into whether the use of speech input reduces participants' time spent on completing the diary entries. To take individual differences into account, we used multilevel linear regression modeling by treating the use of speech input as a fixed effect and participant as a random effect [1]. To further investigate how participants used touch, GS, and LS input, we summarized their data input patterns by grouping the data fields that were typically captured together (e.g., input patterns of capturing task duration, location, and task category in Productivity Diary). We also broke down the usage of the input modalities by each participant to examine

[1]To ensure that the regression analysis would generate valid results, I performed a-priori power analysis for each regression model on G*Power [265]. Based on the number of diary entries collected in Productivity Diary (n = 1032) and Break Diary (n = 382) and corresponding R^2 values, all the models reached over 90% power (α = .05, β = .20) with a medium effect size (*Cohen's* $f^2 > .15$).

the individual variations in modality preferences.

6.3.3.2 Diary Entry Analysis

To examine whether and how input modality influence the data richness of unstructured (i.e., free-form text) input, we analyzed the responses to the three text fields in Productivity Diary (i.e., PD4, PD6, PD7) and the two text fields in Break Diary (i.e., BD2, BD3). First, three researchers independently analyzed a subset of the 3860 text field responses (811, 21%) and created a set of labels characterizing the richness of the responses. Through rounds of comparison and discussions, we agreed to categorize the responses into three categories: *generality*—vague answers lacking details, *specifics*—concrete answers with details, and *specifics with additional contexts*—concrete answers with details and additional contexts that help researchers better understand the situation (See Table 6.4 for details). This categorization follows prior works on analyzing the data richness of open-ended survey responses [17, 233], which was also used in the FoodScrap Study (Chapter 5). We initially coded different types of contexts in the responses (e.g., other people, task procedure, prior work experience), but did not find prominent themes from these contexts. Based on the coding scheme, two researchers revisited the same subset of data and separately coded them, reaching near-perfect (Cohen's κ = .84). After resolving the discrepancies, I coded the remaining responses. Next, we used multinomial logistic regression to examine if input modality tended to affect the data richness of each text field, while treating participant as a random effect.

6.3.3.3 Interview Data Analysis

We audio recorded all the interviews and transcribed them into text. Three researchers separately analyzed the transcripts and built an initial list of codes through rounds of discussions. Focusing on the contexts of how and why participants used touch and speech input, we started with a top-down (deductive) approach to identify factors influencing participants' modality choice. After several iterations of coding, we organized our codes into emerging themes using bottom-up (inductive) thematic analysis [266].

6.4 Results

Over the two weeks, NoteWordy collected 1032 entries in Productivity Diary (60.7 entries per participant) and 382 entries in Break Diary (22.4 entries per participant). As Table 6.2 shows, 43.4% of the diary entries were completed by touch-only input, 12% were completed by speech-only input [2], and the remaining 44.7% were completed with some data fields filled by touch input and others filled by speech input (touch + speech). On average, participants spent 143.7 seconds per entry in Productivity Diary and 78.4 seconds

Input modalities	Total entries	Productivity Diary		Break Diary	
		# of entries	Avg. time spent	# of entries	Avg. time spent
Touch-only input	613 (43.3%)	429 (41.6%)	175.9	184 (48.2%)	86.7
Speech-only input	169 (12.0%)	38 (3.7%)	115.9	131 (34.3%)	65.5
Touch + Speech input	632 (44.7%)	565 (54.7%)	121.1	67 (17.5%)	81.0
Total	1414	1032	143.7	382	78.4

Table 6.2: The number of entries that were completed by touch, speech, and speech plus touch input in the two diaries, together with the average time spent (seconds) on completing the entries.

[2] I use "speech-only input" to denote people using LS or GS input to enter their data, although it requires touching the speech button (i.e., the "push-to-talk" operation).

per entry in Break Diary. We found that speech-only and touch + speech input entries took less time to complete than touch-only entries. The linear multilevel regression modeling showed that the use of speech input significantly reduced the time spent on completing the entries in Productivity Diary (b [3] = -0.38, p = .004). In this section, I present participants' usage of touch and speech input (RQ3) and how the input modality affect the data richness of free-form text fields (RQ4). Additionally, I report the speech recognition and data mismatching issues that participants encountered and how they reacted to these issues.

6.4.1 Usage of Input Modalities

Table 6.3 summarizes participants' input patterns contributing to the data fields that were typically captured together, including a combination of structured (i.e., timespan, multiple choice, Likert scale) and unstructured data (i.e., text). The following consists of findings from three aspects: (1) participants' choice of input modalities for capturing different types of data, (2) their GS usage, and (3) variations in modality preferences across individual participants.

6.4.1.1 Modality Choice By Data Type

Our findings showed that touch input was most frequently used for capturing structured data including timespan, multiple choice, and Likert scale questions (**T1, TS1**), because the interaction was "*convenient*" and "*familiar*" to many of our participants. In the meantime, timespan as a structured data was frequently captured by speech input (GS or LS) as well: in 32.7% of the Productivity Diary entries (**TS1, S1**) and 43.2% of the Break

[3] b refers to regression coefficient.

Data fields	Input pattern	Modality	Freq	Example
Structured data				
[a] **Timespan & multiple choices (*n* = 1032)**	**T1.** T <timespan> + T <multiple choices>	T	660 (63.9%)	Pick start & end time in task duration , tap an option in location and task category
	TS1. GS/LS <timespan> + T <multiple choices>	T GS LS	141 (13.7%)	or task duration say " *8 to 10 a.m.* ," tap an option in location and task category
	S1. GS <timespan & multiple choice>	GS	196 (19.0%)	say " *work task* *at home* *from* *9 to 12 p.m.* "
	Miscellaneous		35 (3.4%)	
Structured data + Unstructured data				
[b] **Timespan & text fields (*n* = 382)**	**T2.** T <timespan> + T<text>	T	184 (48.2%)	Pick start & end time in break duration , type in break activity and break duration
	S2. GS <timespan & text>	GS	165 (43.2%)	say "*I* *walked outside* *from* *4 to 4:30 p.m.* , *feeling* *refreshed because the weather was nice* "
	Miscellaneous		33 (8.6%)	
[c] **Multiple choices & text field (*n* = 1032)**	**T3.** T <multiple choices> + T<text>	T	473 (45.8%)	Tap an option in task category , type in task description
	TS2. T <multiple choices> + LS <text>	T LS	343 (33.2%)	Tap an option in task category , task description say "*writing a report for my class*"
	S3. GS <multiple choice & text>	GS	165 (16.0%)	say " *School-related* *tasks* *on python codes* "
	Miscellaneous		51 (5.0%)	
[d] **Likert scale & text field (*n* = 1032)**	**T4.** T <Likert scale> + T<text>	T	434 (42.1%)	Pick productivity score , type in productivity rationale
	TS3. T <Likert scale> + GS/LS <text>	T GS LS	428 (41.5%)	Pick productivity score , productivity rationale say "*Got most work done fast*"
	S4. GS <Likert scale & text>	GS	149 (14.4%)	say "*I was* *somewhat productive* *because I* *completed the task it but ended up taking more time than planned* "
	Miscellaneous		21 (2.0%)	
Unstructured data				
[e] **Multiple text fields (*n* = 1414)**	**T5.** T <each text field>	T	619 (43.8%)	Type in productivity rationale and feelings respectively
	S5. LS <each text field>	LS	390 (27.6%)	productivity rationale say "*I wasn't very focused,*" feelings say "*tired because I did a lot of chores today*"
	S6. GS <all text fields>	GS	379 (26.8%)	say "*I* *had some snacks* *and* *felt satisfied* *because those are my favorites* "
	Miscellaneous		26 (1.8%)	

[a] Productivity Diary: task duration (PD1) & location (PD2) & task category (PD3).
[b] Break Diary: break duration (BD1) & break activity (BD2) & feelings (BD3).
[c] Productivity Diary: task category (PD3) & task description (PD4).
[d] Productivity Diary: productivity score (PD5) & productivity rationale (PD6).
[e] Productivity Diary: productivity rationale (PD6) & feelings (PD7); Break Diary: break activity (BD2) & feelings (BD3).

Table 6.3: Summary of input patterns contributing to data fields that were typically captured together. The modality column indicates the input modalities that were used (T: touch, GS: Global Speech, LS: Local speech).

Diary entries (**S2**). Some participants found that speech input is more effective than touch input in capturing timespan, because "*manually selecting when it started and ended is tedious, cause you need to select hours and then minutes. So I just went to the task duration mic (LS on task duration) and clicked on it, and say things like '9 to 12 p.m. yesterday' and found it very easy*" (P7). With speech input, participants also described their task and break duration in different ways, such as providing the standard start and end times (e.g., "*8 to 9:30 p.m.*"), referring to relative time points (e.g., "*started 3 hours ago till now*"), or mentioning the duration (e.g., "*started at noon and lasted 45 minutes*"). In addition, participants acknowledged the convenience of capturing free-form text using speech (both LS and GS): "*I probably would never manually input the open ended questions unless I really had to, because it would just take too much time to type the details*"(P14).

6.4.1.2 GS Usage

Starting with GS for multiple data capture: Overall, six participants (P7, P9, P11, P13, P14, P17) showed a strong preference for GS because it was "*faster*," "*intuitive*," and "*more accurate than expected*." Oftentimes, they started capturing everything together and then adjusted individual data fields as needed: "*I started off with the global speech. For the most part, it did a good job capturing what I was saying. Sometimes there will be just spelling errors, so I would make manual adjustments*"(P9). It was noteworthy that although participants rarely used LS to individually capture multiple choice and Likert scale questions, they used GS to capture these two types of data together with other data fields (**S1, S3–4**), because GS saved their effort to "*click and hold for every single field*"

(P10). Likewise, when asked to compare between LS and GS, P7 remarked that LS was like "*individual voice commands*" and GS was more "*close to natural language*."

Transition from LS to GS: Interestingly, P11 and P13 did not use GS at the beginning of the study because they were unfamiliar with the new interaction paradigm: "*Initially, I wasn't really sure about the commands for the global ... So just for my own reliability sake, I was typing it or use the individual ones (LS)*" (P13). But later, they were able to adopt and get comfortable with GS, as P13 explained: "*But I figured it might as well be worth a try to use it. And then you know, after some 'trial and error,' once I had that down, it became just kind of a go-to*."

GS Usage in Productivity Diary vs. Break Diary: In Productivity Diary, GS was used in less than 20% of entries (**S1**, **S3–5**); while in Break diary, GS was used in 43.2% of the entries (**S2**). Participants found GS most useful when they could "*naturally link multiple data in one sentence*", but in Productivity Diary, it was not always as intuitive to do so: "*Sometimes I will try to say things like 'I worked on school-related things at school' or 'working on a work-related task at workplace,' which for me sounds a little awkward to say*" (P11). Nine out of 17 participants explained that they preferred using GS in Break Diary, because the diary was shorter and all the data fields were visible on the screen at once, allowing them to quickly skim what information to capture and speak without scrolling: "*I could see everything on the screen at same time, so I didn't have to worry that I was going to miss a question or something like that*" (P14).

Adoption barriers of GS: Six participants (P1, P2, P3, P4, P6, P12) rarely used GS during the study, partly due to their unfamiliarity, but also because including multiple data

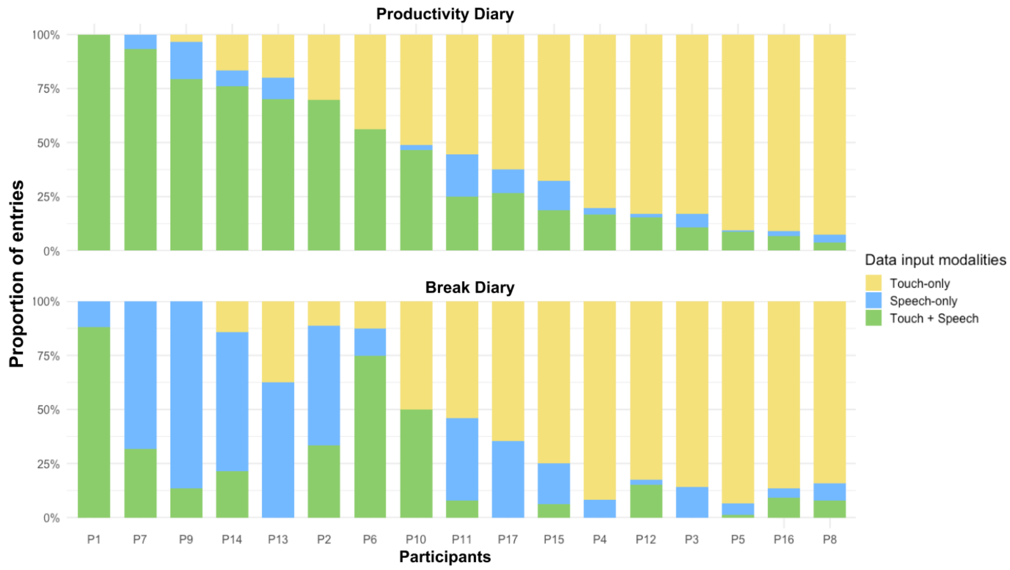


Figure 6.6: The usage of speech and touch input per participant in completing the diary entries (participant id is ordered by the proportion of entries involving speech input in Productivity Diary).

fields in one sentence could take extra mental load. For example, P3 noted that "*although the global speech was really really cool, I found myself not ready to use it ... I didn't always have all my thoughts together of exactly what I wanted to say for every single part. I would forget what else needed to be said so I'd have to stop and think.*" Sometimes, participants provided long responses to describe their tasks, productivity rationale, and feelings. In these cases, they preferred LS rather than GS for capturing details in each field: "*I always provide as many details as possible for the study. If I use the global, I would add more details with the local mic (LS) anyway, so I did not use it very often*" (P6).

6.4.1.3 Variations in Modality Preferences

We noticed large variations in participants' modality preferences. As Figure 6.6 shows, eight participants (P1, P7, P9, P14, P13, P2, P6, P10) used speech input in more than 50% of their diary entries; three participants (P11, P17, P15) used speech input occasionally, but less often than touch input; and the remaining six participants (P4, P12, P3, P5, P16, P8) used touch input most of time, with fewer than 25% of diary entries involving speech input. During the interviews, those who were already comfortable with speech interfaces expressed their excitement about the ***convenience*** and ***accuracy*** of speech input: "*I would say it's pretty accurate. The global speech is very impressive, you don't really need to remember the keywords specifically, because as long as you follow the diary, it catches what you are trying to say*" (P7). They also enjoyed "*thinking out loud*" with speech input, because it made them "*feel accomplished about the tasks*" (P2).

Despite the advantages, participants were concerned about using speech input to capture their data for several reasons. First, due to ***privacy concerns***, they did not feel comfortable talking about their tasks, productivity, and breaks when other people were around (P6, P11, P12, P15, P16, P17), especially in office settings. For example, P12 noted that " *I don't want anybody, like my colleagues or my boss to hear that I just took a break from doing work and was on screen all the time*." Second, participants pointed out that capturing personal information by talking to their phones was ***not a "social norm"*** in public spaces (P3, P4, P5, P10, P12). Rather than worrying about privacy, they felt embarrassed about "*over-sharing*" their life that others did not care about. Third, some participants found themselves "***better at writing than speaking***" (P3, P5, P8, P16, P17). In the cases where they needed to describe complicated thoughts and emotions, they preferred manually typing the information, as P17 mentioned: "*I think the rationale of the productivity score and the feeling about the tasks had a little more involvement. So I guess for me, it's just easier to write that out than speaking out*."

6.4.2 Data Richness

While analyzing the data richness of responses to free-form text fields, we grouped each response into one of the three categories: *generality*, *specifics*, and *specifics with additional contexts*. We note that *specifics with additional contexts* were responses that already provided specific details, along with *additional contexts* which can be removed without affecting the completeness of the response [17, 233, 249]. In Break Diary, we found that most of the responses in Break Diary were under the specifics category (BD2: 96.3%; BD3: 91.4%), suggesting a small variation in data richness. Thus, we focused on examining the responses in Productivity Diary.

Table 6.4 describes how we characterized data richness of each text field with examples, and table 6.5 summarizes the number of responses in each category and the input modalities involved. *Responses involving speech input* include responses that entered by speech-only input as well as those that entered by both touch and speech input. We did not differentiate these two types of responses because (1) a majority of the responses entered by both modalities were captured by speech input and slightly edited by touch input for spelling or punctuation issues; (2) the number of these responses only took a small proportion (PD4: 9.3%, PD6: 8.9%, PD7: 8.5%). We noticed that 51 (1.6%) responses in Productivity Diary digressed from the original questions (e.g., answering "*My family felt happy because they had missed me*" to feelings and why (PD7)) and excluded them from the regression analysis.

The logistic multilevel regression modeling showed that input modality tended to affect the data richness of all three text fields: task description (R^2 = .15, p < .001), pro-

Text field	Generality	Specifics	Specifics with additional contexts
Task description (PD4)	General task type lacking details "*had a meeting*" "*coding*"	Specific about task activities "*I edited two video clips for our YouTube channel*" "*Writing python code for my class*"	Specific about the task with additional contexts other than time and location asked in the diary (e.g., colleagues, procedure) "*I attended a UX meeting with other designers. We shared some case studies applying design thinking and talked to the BA team for next steps*"
Productivity rationale (PD6)	Vague about the productivity rationale "*It's not the most productive time*" "*Productivity same as before*"	Rationale clearly explaining why they were productive or not "*Because I got everything done in the time expected without distraction*"	Clearly explained why they were productive or not and elaborated the response with additional contexts (e.g., task outcome, upcoming events) "*This meeting went really well and we had a great discussion. There were no instances of unresolved questions or topics in preparation for our Thursday morning meeting*"
Feelings (PD7)	Vague about why they felt in certain ways "*I felt challenged and*" "*frustrated*"	Specific reasons explaining how they felt and why "*I felt great during the task because I was caffeinated enough and I had a good conversation with my student*"	Clearly explained how they felt and why and elaborated the response with additional contexts (e.g., emotion fluctuation, long-term plans) "*I felt discouraged at first because I didn't know what I would write about, then I felt inspired because I found a theme. Then I felt really happy because I was able to submit the assignment. Overall I felt proud for completing a task I had considered skipping and believe I did a good job*"

Table 6.4: The definitions of data richness categorization of each text field in Productivity Diary, together with examples from diary entries. In particular, *specifics with additional contexts* were responses that already provided specific details, along with *additional contexts* which can be removed without affecting the completeness of the response. In the examples, we highlighted those additional contexts in blue.

Text field	Responses with touch-only input				Responses involving speech input			
	Total	Generality	Specificity	Specificity with additional contexts	**Total**	Generality	Specificity	Specificity with additional contexts
Task description (PD4)	**479**	202 (42.2%)	205 (42.8%)	72 (15.0%)	**518**	102 (19.7%)	306 (59.1%)	110 (21.2%)
Productivity rationale (PD6)	**614**	151 (24.6%)	394 (64.2%)	69 (11.2%)	**410**	52 (12.7%)	254 (61.9%)	104 (25.4%)
Feelings (PD7)	**588**	198 (33.7%)	352 (59.9%)	38 (6.4%)	**436**	134 (30.7%)	228 (52.3%)	74 (17.0%)
Total	**1681**	551 (32.8%)	951 (56.6%)	179 (10.6%)	**1364**	288 (21.1%)	788 (57.8%)	288 (21.1%)

Table 6.5: Input modalities (responses entered by touch-only input versus responses involving speech input) x data richness (generality, specificity, and specificity with additional contexts) for each text field in Productivity Diary. Note that we excluded responses that digressed from the questions.

ductivity rationale (R^2 = .15, $p < .001$), and feelings (R^2 = .10, $p < .001$). In task description (PD4) and productivity rationale (PD6), responses involving speech input were more likely to be specific (PD4: *OR* [4] = 3.79, $p < .001$; PD6: *OR* = 2.16, p = .002) and include additional contexts (PD4: *OR* = 3.0, $p < .001$; PD6: *OR* = 4.18, $p < .001$). In feelings (PD7), although responses involving speech input were not necessarily more likely to be specific (*OR* = 1.20, p = .36), they tended to include additional contexts (*OR* = 2.12, p = .03). These findings were corroborated during the interviews, as participants recalled that with speech input, they were inclined to enter more details and express their thoughts more freely (P1, P10, P11, P13, P15, P17): "*In a natural way, I definitely put more using speech, because I can just talk, and typing is more time consuming. Like speech is a more free and natural way for me to express my thoughts, I guess especially for productivity (rationale) and how I felt*" (P1).

[4] *OR* refers to odds ratio.

6.4.3 Reactions to Speech Recognition & Data Mismatching Issues

Although participants generally had a positive experience with speech input, we identified several speech recognition and data mismatching issues that influenced their experience. Among the 103 invalid utterances (from both LS and GS input) that were logged by NoteWordy, 81 of them intended to fill specific data fields but failed for the first time (see Table 6.6), 10 intended to fill specific data fields after the first attempt but failed again, and 12 were unspecified. In what follows, we elaborate on these cases.

6.4.3.1 Timespan & Likert Scale Questions

Most of the recognition issues were related to timespan and Likert scale questions, which occurred when the system misrecognized the words that are part of the data value. For example, some numbers when spoken out loud (e.g., "*two*," "*four*," "*five*") could be misrecognizeed as different words with similar pronunciations (e.g., "*to*," "*for*," "*fine*"), or vice versa. As a result, utterances such as "*started 5 and lasted 4 hours*" could be recognized as "*started fine and lasted for hours*." Among the 66 invalid utterances that intended to fill task and break duration, we found 42 (63.6%) second attempts that participants made to enter the data using speech input either via LS or GS. Participants acknowledged that when their speech input for task and break duration was not recognized, they were willing to make one more attempt with speech input, "*which usually worked*" (P2). In the case where the duration was partially misrecognized (i.e., incorrect start or end time), however, participants tended to correct the issue using touch input: "*I would just manually fix the error because I only needed to do it for one time point rather than repeating both*" (P9).

Among the 11 invalid utterances for productivity score, we found only one second attempt using speech input.

6.4.3.2 Multiple Choice Questions

In a few cases, recognition issues occurred in multiple choice questions when the input included words beyond the recognizer's vocabulary. For example, the system could not recognize the location information from the utterance "*working in the lab*" without words related to "*school*," "*workplace*," or "*others*." When participants encountered such issues, they never made additional attempts with speech input because "*it obviously took more time to click the individual microphone (LS) and say 'school' than simply selecting 'school' on the screen*" (P9).

6.4.3.3 Other Recognition Issues

We note that the speech recognition issues in Table 6.6 were not exhaustive, due to the lack of ground truth to identify misinterpreted information or spelling errors in text fields. For example, in responding to productivity score, utterances such as "*moderately productive*" or "*relatively productive*" were interpreted as "*productive*" (*6*), even though they were more likely to indicate "*somewhat productive*" (*5*). Another common issue was

Data fields	**Total #**	**# of second attempt with speech input**	**# of additional attempts with speech input (*median*)**	**# of additional attempts with speech input (*max*)**	**# of instances fixed by touch input**
Task/break duration	66	42 (63.6%)	1	4	26 (39.4%)
Productivity score	11	1 (9%)	0	1	10 (91%)
Location/task category	4	0	0	0	4 (100%)

Table 6.6: Instances of speech recognition issues related to different data fields and how participants reacted to them.

related to punctuation in text fields: if participants paused while speaking, the system would add a punctuation mark (e.g., period) right after the current input, which was not always the correct place. Three participants (P3, P8, P16) reported that this issue led them to avoid speech input for entering long sentences: "*I really preferred the manual typing, because I wrote really long sentences, which the speech recognition couldn't get all of it, it kept interpreting my pauses as periods when they should have been commas*" (P8).

6.4.3.4 Strategies to Avoid Recognition Issues

From the experience, some participants learned to avoid similar issues by intentionally selecting the words in their utterances. For example, some participants (P7, P14, P17) found that mentioning time-related phrases such as "*a.m.*" and "*in the morning*" could improve the accuracy of time recognition. After realizing that the word "*to*" was often recognized to the number "*two*," P15 decided to use the word "*till*" to describe timespan (e.g., "*7 till 9*"). In describing their productivity score, participants used the label (102, 68.5%) more often than the number (47, 31.5%) (e.g., saying "*very productive*" instead of "*7*"). P10 and P17 explained that this was the strategy they chose to prevent the system from missing the productivity score. However, other participants had lower tolerance for such recognition issues, and therefore turned to touch input most of time: "*because of my accent I guess, it didn't work very well in the first place, so I just feel more comfortable with touch or writing*" (P16).

6.4.3.5 Data Mismatching Issues

When it comes to capturing multiple data fields using GS, the system could mismatch participants' utterances to irrelevant text fields. In this case, participants could delete the text by tapping the clear button . During the study, we found that participants used the clear button 110 times. After examining the transcribed utterances right before the use of clear button, we identified 77 text mismatching instances. Among these instances, 49 (63.6%) were due to the filler words at the beginning or end of the utterance. For example, the utterance "*From 9 to 11 a.m., I was doing work-related tasks at home. Yeah*" only intended to capture task duration (PD1), location (PD2), and task category (PD3) in Productivity Diary. However, the filler word "*Yeah*" was placed into task description (PD4) because of the way that NoteWordy handles uncategorized text. The other 28 (36.4%) instances all occurred in Productivity Diary, where participants intended to capture their productivity score (PD5) by including the word "*feel*" or its other forms (e.g., "*I felt productive*"). Although the productivity score was correctly recognized, the text segment also appeared in feelings (PD7). When data mismatching occurred, all the participants agreed that the clear button was helpful to delete the text and reenter the correct information with either LS or touch input.

6.5 Discussions

The quantitative results on data input patterns showed how participants used touch and speech input to capture different types of data; the qualitative findings from the interviews further explained the advantages and limitations of the two input modalities. With the lessons learned, I discuss opportunities for effectively combining touch and speech input to support data capture in various self-tracking contexts.

6.5.1 Integrating Touch & Speech to Support Capturing Different Types of Data

We found ***speech*** input significantly reduced the time spent on completing the diary entries and helped enhance data richness of free-form text. These findings echoed with prior study on speech-based food journaling, suggesting that speech input was perceived easy to use and could encourage people to elaborate their responses from different aspects [249]. Thus, speech input holds promises to lower the data capture burden while collecting rich contexts, which is important for collecting self-reported behaviors and assessments—data that are difficult to be automatically captured or to interpret due to lack of contextual information (e.g., health symptoms [218], mood [27], reflective thoughts [267]). ***Touch*** input was frequently used for capturing structured data including timespan, multiple choice, and Likert scale questions, especially the latter two that require only a single tap. Participants were also more comfortable with touch input in public spaces, where they concerned about privacy or social appropriateness. Even for the same data field, participants' modality preferences might differ depending on the complexity and sensitivity of the data (e.g., not wanting colleagues to hear about their non-productive tasks). As such, our study demonstrated the effectiveness of integrating both touch and speech input to support capturing multiple types of data in different scenarios.

We noticed that even though participants rarely used LS to individually capture multiple choice and Likert scale, they still used GS to capture these data fields together with other data. This observation prompted us to rethink how to make the best use of speech input while incorporating it into an existing system. A design consideration is to provide speech input based on the composition of the data capture regimen. For example, data fields that can be filled with a single tap might not need LS, but GS can be helpful for collectively capturing these data fields if they naturally belong in the same utterance. In section 6.5.3, we discuss how to support efficient multi-data capture in more detail.

6.5.2 Enabling Flexible Time Capture & Editing

Time is an important component in self-tracking (e.g., sleep duration [11], eating time [249]), but is laborious to capture on smartphones: people often need to manually pick the date, hour, minute, or range. In our study, participants used both touch and speech input to capture their task and break duration. Compared with touch input (i.e., manually picking the start and end time), they acknowledged the flexibility of speech input to describe timespan in different ways. Even when speech recognition issues occurred, participants were generally willing to enter the timespan data by making a second attempt with speech input. However, if only one of the time points needs to be updated, participants preferred touch input. This was partly due to NoteWordy's limitation in processing timespan data, which required people to provide both start and end time (or duration) at once. To support flexible timespan editing, we can enable people to select the target time point and then update it with speech input (e.g., pressing on the end time field and say "*9 p.m.*") or specify the component that they intend to update in their utterances (e.g., "*end time: 9* *p.m.*,” “*duration: 30 minutes.*”).

6.5.3 Supporting Efficient Multi-Data Capture With Speech Input

Our study generated several input patterns of using GS to capture multiple data fields, including combinations of different structured and unstructured data. Participants acknowledged that GS was fast and intuitive, especially when they were able to naturally link the data fields together. But sometimes it could be unnatural or redundant to include multiple data fields in one sentence (e.g., “*working on a work-related task at workplace*”). In addition, participants preferred using GS for Break Diary more than Productivity Diary, because all the data fields in Break Diary were on the screen at once. To support efficient multi-data capture with speech input, one design opportunity is to display semantically-related data fields on the same screen, so that people can easily skim what data to capture and then naturally phrase their utterances. However, identifying the semantic relationships between multiple data fields can be challenging, especially when the number of data fields increases [268]. Therefore, more sophisticated techniques (e.g., machine learning [269]) are needed to predict the possible linguistic structures to include multiple data fields in a sentence, so that researchers can better design their data capture regimens.

Another challenge of adopting GS was unfamiliarity. Participants mentioned that they were unsure about how to properly phrase an utterance containing multiple data fields or felt mentally taxing to come up with such an utterance. Even among those who were able to figure out how to use GS, some felt the similar uncertainty at the beginning of the study. Although we provided preemptive guides in NoteWordy (see Figure 6.4b) and conducted practice sessions with GS during the tutorial, these were not enough to overcome the adoption barriers. We suspect that people need more "prompts" to get used to GS, as P13 indicated, he was able to overcome the unfamiliarity after some "*trial and error*."

6.5.4 Adapting Speech Recognizers for Self-Tracking Activities

Like many speech input systems, NoteWordy embeded a commercial API to recognize speech into text. However, these commercial speech recognizers are trained with context-agnostic dataset and are not fine-tuned for self-tracking activities. Not taking the personal data capture context into account, the speech recognizer that we embedded sometimes failed to handle ambiguity in word pronunciations. Hence, data fields such as timespan and Likert scale are particularly vulnerable to speech recognition errors: if one of the keywords was misrecognized, the entire utterance could became invalid (e.g., "*7 to 9*" being recognized as "*729*").

Although speech recognition services such as Microsoft Cognitive Speech API [270] and Google Cloud Speech-to-text [271] allow developers to upload their own training dataset for customized recognition, they often require these dataset to be large enough to cover multiple speech variances (e.g., accents, dialects), which is not yet available in the domain of self-tracking. To better adapt the speech recognizers for personal data capture, we call for large-scale research efforts to generate contextualized training data from diverse self-tracking activities, including but not limited to date and time [16], commonly used labels of Likert scale (e.g., sleep quality [11], stress level [272]), and units for describing daily activities (e.g., cups of coffee [11], exercise repetitions [91]).

6.6 Chapter 6 Summary

In this chapter, I reported a two-week long data collection study with NoteWordy, a multimodal mobile app integrating touch and speech input to capture different types of data, in the context of productivity tracking. During the study, 17 working graduate students collected data about their tasks and breaks, and generated several input patterns with touch and speech input. The study demonstrated how data types interplaying with participants' input habits, error tolerance, and social surroundings contributed to the ways that they used touch and speech input. Additionally, I examined how speech input affected the diary completion time and data richness of unstructured input. With the lessons learned, I discuss implications for leveraging the strengths of touch and speech input to better help people collect different types of personal data and to improve the data capture experience with natural language input. I also discuss opportunities for adapting speech recognizers for other self-tracking activities outside the study context.

Chapter 7: TandemTrack: Examining How a Smart Speaker Complements a Mobile App in Supporting Exercise Training and Tracking

In Chapters 5 and 6, I focused on examining how multimodal input can support data capture on mobile phones and highlighted the promise of speech input in lowering the data capture burden, enhancing data richness, and promoting situated reflection. In this chapter, I expand the research scope by introducing smart speakers as self-tracking devices, aiming to answer **RQ5**: How does a smart speaker complement and augment a mobile app in supporting consistent exercise? I built a multimodal system TandemTrack, coupling a mobile app and an Alexa skill on Amazon Echo devices to support in-home exercise training and tracking. Through a four-week between-subjects study with TandemTrack, I report participants' exercise adherence and performances, along with their preferences for the app and the skill. The findings led to discussions on how a smart speaker can complement a fitness mobile app in supporting exercise tracking, and how we can design better multimodal self-tracking systems across devices.

7.1 Introduction

Technologies such as wearable devices and mobile apps have been developed to assist exercise, supporting workout planning and tracking, as well as providing progress feedback. While these technologies can help people achieve their exercise goals, a large amount of population that could benefit from these technologies still does not engage with them [138], failing to meet the recommended level of exercise guidelines [143].

Smart speakers present promising opportunities for supporting consistent exercise in the home environment: (1) the hands-free interaction can lower the data capture burden during exercise; (2) in the home environment where people do not carry their mobile phone all the time, the voice reminder from the smart speaker is more noticeable than a regular mobile notification; (3) smart speakers' lack of mobility can force a consistent exercise location, which is helpful for habit formation [273]. However, most of the smart speakers in the market do not provide sufficient visual feedback, which is a valuable medium for promoting reflection in self-tracking. Popular fitness apps, on the other hand, have rich visual elements while neglecting the convenience that the smart speaker may provide.

In this light, my colleagues and I examine how a smart speaker's voice interaction complements and augments a mobile app in supporting consistent exercise. We designed and developed TandemTrack, a multimodal system coupling a mobile app and an Alexa skill, which offers a simple exercise regimen alternating between sit-ups and push-ups, captures workout repetitions, provides performance feedback, and sends daily reminders. The two modalities share the same database in the cloud, supporting quick data synchronization between the app and the skill, so that people can use them interchangeably.

Through a four-week between-subjects study by deploying TandemTrack to 22 participants, we collected rich data on participants' exercise adherence and performance, and their preferences for the app and skill, while examining how TandemTrack as a whole influenced their exercise experience. With the lessons learned, I discuss the opportunities and challenges for speech interaction to assist daily exercise, and implications for designing multimodal system to promote consistent exercise routines.

7.2 TandemTrack Design and Implementation

The goal of this research is to maximize the advantages of the smart speaker and the mobile app, instead of competing one against the other. Therefore, we designed TandemTrack as a research prototype that situates people to interact with a smart speaker and a mobile phone in a simple exercise context, not as the most powerful exercise intervention. In the following, we describe the the design goals of TandemTrack, key components, and implementation details.

7.2.1 Design Goals

The first design goal (DG1) is to **lower exercise barriers**. According to prior research, common barriers to performing consistent exercise include lack of time, inconvenience, and environment constraints [138–142]. To reduce the workout complexity and time spent, we incorporated a simple and basic exercise regimen that allows people to complete the exercise session quickly at any location. Specifically, we choose sit-ups and push-ups, two common indoor exercise routines that can be completed with minimal learning efforts, but also provide sufficient health benefits [274,275].

The second design goal (DG2) is to **leverage smart speakers' unique characteristics**. The Amazon Echo provides a set of features, some of which overlap with those of mobile devices, while others are unique. Because the overarching goal is to leverage the synergy that come from both interaction modalities, TandemTrack was designed to explore what modality would better suit people's needs and preferences in delivering exercise regimen, supporting data capture, providing feedback, and sending reminders.

The third deisgn goal (DG3) is to **promote consistency in exercise time and location**. Behaviors become automatic after they have been performed consistently and repeatedly [276]. The key is to performing the target behavior every day, especially at the same time and place [276]. Therefore, we aim to encourage exercise on a daily basis, ideally taking place at the same time and the same place.

7.2.2 TandemTrack Design Components

The two modalities share the same database and support the same key features: an exercise regimen alternating between sit-ups and push-ups, data capture for exercise performance, exercise feedback, and daily reminders. Here, I describe how we achieved the above design goals with the four key components.

7.2.2.1 A Simple yet Beneficial Exercise Regimen

Instead of providing diverse and complicated workout guidance like existing fitness apps, TandemTrack minimizes the workout complexity to allow people to start exercising easily (DG1). TandemTrack delivers an exercise regimen that alternates between sit-ups and push-ups by day, focusing on timing and repetitions (DG3). The entire exercise session on TandemTrack lasts for three minutes and 30 seconds, including three exercise sets (either sit-up or push-up) and two breaks in between. Each set lasts for 30 seconds and each break lasts for 60 seconds. Figure 7.1 shows an example exercise flow of using the TandemTrack skill to do push-ups. To allow people to control their exercise pace, TandemTrack provides options for skipping the break or resetting the current set.

7.2.2.2 Voice-Activated Data Capture

Exercise data can be captured at different dimensions, ranging from a simple record to detailed metrics. For example, RunKeeper [277] measures runners' performance by capturing their running distance, average pace, time spent, path, heart rate, calorie burned, etc. JEFIT [278], on the other hand, enables people to manually capture exercise-related information in different formats such as number of repetitions, text notes, body metrics (e.g., weight), and photos.

TandemTrack lowers the data capture burden (DG1) by capturing only the number of repetitions per set for sit-up/push-up right after the completion of each set when the number is still fresh in people's mind. In the mobile app, a pop-up a text field asks for people to manually enter their repetitions, while the Alexa skill prompts the user to speak out their sit-up or push-up repetitions by asking *"how many sit-ups/push-ups did you complete?"*, and takes their speech responses as input. When all three sets are complete, the TandemTrack skill records individuals' exercise data and sends a confirmation message:

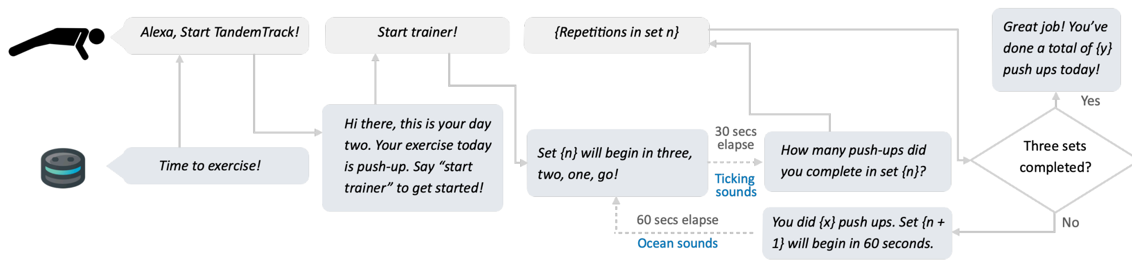


Figure 7.1: An example of using the TandemTrack skill to do push-ups.

"*Great job! You've done a total of X sit-ups/push-ups today*." Unlike most fitness training skills that do not support data capture or only provide feedback on the data captured by other devices, the TandemTrack skill supports data capture independently of other devices. In case of typos on the mobile app or Alexa's inaccurate speech recognition, TandemTrack allows people to edit their repetition data after exercise in the mobile app.

7.2.2.3 Feedback With Multimodal Interaction

With the exercise data collected, the TandemTrack app provides visual feedback in three parts (Figure 7.2): (1) the daily exercise feedback on the top (A) shows the sit-up/push-up repetitions; people can swipe left and right to review the repetitions of sit-ups or push-ups they completed on each day; (2) the middle part (B) summarizes the longest streak (i.e., the consecutive days of doing exercise) and the total number of completed exercise sessions; (3) the aggregated feedback on the bottom shows a series of time-based visualizations, including exercise streak view (C1), sit-up performance (C2), and push-up performance (C3).

Complementing the visual feedback, TandemTrack skill provides auditory feedback in a Q&A manner (DG2). People can ask about their "*exercise summary*" to receive a sum-

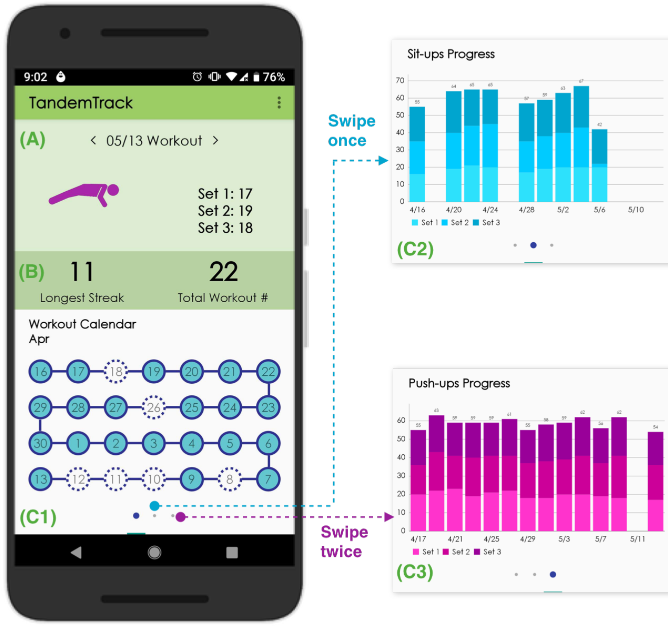


Figure 7.2: The home screen of the TandemTrack mobile app: the daily exercise feedback (A); a summary of the longest streak and complete exercise sessions (B); a series of aggregated feedback—exercise streak view (C1), sit-up progress (C2), and push-up progress (C3).

mary of their exercise progress, including total exercise sessions, current streak, longest streak, and average sit-up/push-up repetitions per session. They can also ask specific questions on their exercise records, such as "*how many push-ups did I complete yesterday*" and "*what is my best sit-up performance*."

7.2.2.4 Daily Reminders to Facilitate Exercise Routine

TandemTrack emphasizes on the importance of consistent routines by suggesting a daily exercise. Because forgetfulness is one of the most common reasons that prevent people from building a habit [279], TandemTrack asks people to set a daily reminder by picking a time when they want to exercise (DG3). Considering that people may not be close to the Amazon Echo when the reminder is sent, both modalities of TandemTrack remind people with the same message, "*Time to exercise*." The reminder of the mobile app comes from a regular notification, and the reminder of the Alexa skill comes from the Echo device's internal voice reminder. To encourage a consistent exercise routine, TandemTrack does not allow people to change their reminder time.

7.2.3 Implementation

The TandemTrack mobile app was implemented in Android Studio [280] using Kotlin [261]. For the TandemTrack Alexa skill, we used Alexa Skills Kit (ASK) [281] to build the front-end voice interface, and Node.js [282] for the back-end service, which is hosted on the Amazon Web Service Lambda [283]. The TandemTrack skill is distributed through Amazon Alexa skill's beta test, which is only available to people who are invited to test the skill.

The TandemTrack app and the skill share access to the same database on Google Firebase [284], which synchronizes people's exercise data on both devices. TandemTrack allows people to hear their exercise feedback from the skill while reviewing the visual feedback on their phone, but it does not support people to perform exercise with the app and the skill at the same time due to the technical difficulties.

To explore what modality would better suit people's needs and preferences in delivering exercise regimen, supporting data capture, providing feedback, and sending reminders, we strove to provide equivalent features for both the mobile app and Alexa skill to ensure that they have the same capabilities. In this way, people could use the app and the skill without being influenced by extra or missing features on either modality, and we could examine the specific advantages and drawbacks of voice interaction (on smart speakers) regarding each feature. However, the mobile phone and the Amazon Echo has its own characteristics, making some of the interaction experience inherently different. For example, the mobile notification stays on the screen as long as the user does not remove it; but the Alexa reminder sent from the Amazon Echo does not have the same visibility due to the ephemeral nature of auditory information.

7.3 Methods

We conducted a four-week between-subjects study deploying TandemTrack with 22 participants, followed by semi-structured interviews. All the participants had not have the habit of performing sit-ups and push-ups regularly. During the study, 11 participants used only the mobile app to perform exercise (which we refer to as the "M" group), while the other 11 interacted with both the mobile app and Alexa skill (the "MA" group).

7.3.1 Participants

After the study was approved by the Institutional Review Board (IRB # 1132164-9), we advertised the study through the university mailing list. Among the 44 people who responded, 25 met the inclusion criteria: individuals who (1) are over 18; (2) own an Android phone that runs an operating system 6.0.0 or above; (3) have stable Internet access at home; (4) are motivated to do short strength exercise (i.e., sit-ups and push-ups) daily and collect their exercise data; and (5) currently do not do sit-ups or push-ups regularly (i.e., two times or more per week).

With an aim of creating a balanced group assignment regarding participants' gender, age, living environment, and motivation levels[1], we assigned 12 participants to use the TandemTrack mobile app only (M) and the other 13 to use both the mobile app and the Alexa skill (MA). To ensure that all the MA participants were first-time smart speaker users, I intentionally assigned the three participants who already owned a smart speaker (e.g., Amazon Echo, Google Home) to the M group (M-4, M-8, M-9)[2].

Among the 25 participants who participated in the study, one MA participant withdrew from the study because his phone was broken. During data analysis, we excluded the data of one M participant and one MA participant who were found to be housemates and overlapped for two weeks participating in the study. We thus analyzed the data of the remaining 22 participants (11 in M, 11 in MA; 11 female, 11 male). They were 16 graduate students, 3 undergraduate students, and 3 full-time university staff, and their ages ranged from 20 to 61 (M = 26.0, SD = 8.3). Five M participants and five MA participants were non-native English speakers. All participants lived off-campus with housemates or family members during the study period, except for M-11 who lived independently in an apartment. Participants' demographic details can be found in Table 7.1.

7.3.2 Study Procedure

The study consisted of three stages, including study tutorial, four-week deployment with TandemTrack, and debriefing interview. At the end of the study, each participant received a $60 Amazon gift card as compensation. My research team loaned each MA

[1] I asked participants "*How motivated are you to form a strength exercise habit (i.e., push-ups, sit-ups)?*" in the screening questionnaire to capture their motivation levels using a five-point scale (from 1: "*Not motivated at all*" to 5: "*Very motivated.*")

[2] I use M-# and MA-# to denote a participant ID in two groups.

ID	Age	Gender	Occupation	Motivation Level	Native Language	Additional Household Members
M-1	27	M	Graduate Student	Very motivated	Chinese	3 Housemates
M-2	25	M	Graduate student	Neutral	English	1 Housemate
M-3	25	F	Graduate student	Somewhat motivated	English	1 Housemate
M-4	26	F	Graduate student	Neutral	Indian	A partner
M-5	21	M	Undergraduate student	Very motivated	Indian	1 Housemate
M-6	28	F	Graduate student	Somewhat motivated	Indian	A partner
M-7	22	M	Undergraduate student	Very motivated	English	Parents
M-8	25	F	Graduate student	Somewhat motivated	Indian	1 Housemate
M-9	25	M	Graduate student	Neutral	Indian	1 Housemate
M-10	24	F	Graduate student	Somewhat motivated	Indian	1 Housemate
M-11	61	F	University Staff	Somewhat motivated	English	NA
MA-1	30	M	University Staff	Neutral	English	A partner
MA-2	23	M	Graduate student	Somewhat motivated	English	1 Housemate
MA-3	32	F	University Staff	Very motivated	English	A partner and a child
MA-4	20	F	Undergraduate student	Somewhat motivated	English	Parents
MA-5	32	M	Graduate student	Somewhat motivated	English	1 Housemate
MA-6	29	F	Graduate student	Very motivated	English	A partner
MA-7	27	F	Graduate student	Neutral	Chinese	3 Housemates
MA-8	24	M	Graduate student	Very motivated	Indian	1 Housemate
MA-9	31	M	Graduate student	Very motivated	Indian	3 Housemates
MA-10	26	F	Graduate student	Somewhat motivated	Chinese	1 Housemate
MA-11	25	M	Graduate student	Somewhat motivated	Indian	3 Housemates

Table 7.1: Demographics of the 22 participants in the four-week between-subjects study with TandemTrack.

participant an Amazon Echo Dot (3rd generation), which they returned back after completing the study.

7.3.2.1 Study Tutorial

We first scheduled an in-person tutorial with each participant at a research lab. During the tutorial, we introduced the study procedures, instructed participants to install the TandemTrack mobile app, and helped them set a reminder. We explained that the reminder time could not be changed throughout the study, and asked participants to carefully pick a time when they were most likely to exercise. To demonstrate how TandemTrack works, participants needed to perform a full session of sit-ups by following the TandemTrack app's exercise regimen, on the Yoga mat that we prepared. To make sure that participants understand how to interpret the visual feedback within the app, we explained each part of the visualization in detail.

With MA participants, we created a new Amazon account for each of them, and showed them how to connect the Echo Dot to Wifi. Given that all the MA participants were new to smart speakers, we also demonstrated how to communicate with Alexa using basic speech queries (e.g., asking weather, playing music). We then showed them how to perform exercise and ask feedback with the TandemTrack skill. To ensure that MA participants would not receive duplicated exercise reminders from their Alexa app (i.e., the companion app of Amazon Echo), we asked them to turn off the notification option for the Alexa app (they would still receive the exercise reminder from the TandemTrack app and the Echo Dot). In addition, we clarified that TandemTrack is a multimodal system—

their exercise would be synchronized on both devices, and MA participants needed to put the Echo Dot at a place where they could do exercise nearby. As such, MA participants were encouraged to explore the app and the skill as they liked.

The tutorial lasted about 30 minutes for M participants, and 60 minutes for MA participants. At the end of the tutorial, each participant received a study manual (including a physical copy and a digital copy) that described common usages of the TandemTrack app. The manual that MA participants received also included usages of the Echo Dot and the TandemTrack skill. At the end of that day, we confirmed with MA participants that they have connected the Echo Dot to Wifi.

7.3.2.2 A Four-week Between-Subjects Study

The day after the study tutorial, participants started using TandemTrack to exercise for the following four weeks. At the end of each week, we sent each participant a weekly diary using Google form, asking them to briefly respond to three short questions: (1) "*What was your experience like with TandemTrack over the past week*?" (2) "*Did you experience any technical difficulties with TandemTrack*?" and (3) "*Is there anything you want to share with us*?" The purpose of sending a weekly diary was to collect participants' feedback on TandemTrack as the study progressed, and to check if they encountered any technical issues. We did not send a weekly diary at the end of the forth week, because we were able to talk with participants during the debriefing interviews.

7.3.2.3 Debriefing Interview

We conducted a semi-structured interview with each participants at the end of the study. To help participants better recall their exercise experience, we asked them to refer their exercise feedback on the TandemTrack app while having them describe their exercise locations, environments, and reasons for missing the exercise on certain days. We also asked how they used the four key features (i.e., exercise regimen, data capture, feedback, and reminder) provided by TandemTrack. For MA participants, we asked additional questions regarding their preferences and use of the mobile app and the Alexa skill, and their experience in interacting with the Echo Dot. Each interview lasted 20 to 45 minutes.

7.3.3 Data Analysis

7.3.3.1 TandemTrack App & Skill Usage Analysis

We first compared the two groups regarding their exercise adherence (i.e., complete sessions, longest and average streak) and interaction with the TandemTrack app (e.g., opening the app, reviewing feedback) using an independent *t*-test. Then we examined how participants' exercise performance (i.e., average repetitions per session) changed as the study progressed using a linear regression analysis. We also examined the factors affecting participants' exercise adherence using a multiple linear regression, as well as the factors affecting their performance change using a multiple logistic regression. We considered participants' reminder time as the time when they were expected to exercise, and calculated the *time difference* between their reminders and their actual exercise time.

To examine whether the choice of exercise device affected that time difference, we used multilevel linear regression modeling by treating exercise device as a fixed effect and participant as a random effect.

7.3.3.2 Interviews and Weekly Diaries Analysis

We audio-recorded all the interviews, took verbatim notes, and digitized all the weekly diaries. We analyzed the qualitative data from interviews and diaries together to answer the following questions: (1) why participants missed exercise; (2) what they liked or disliked about the four key features provided by TandemTrack (exercise regimen, data capture, feedback, reminder); (3) what factors influenced MA participants' choice of the modality in interacting with TandemTrack; and (4) what challenges participants encountered when interacting with TandemTrack.

We used an inductive approach to first identify codes as labels from participants' elaborations pertaining to the above questions. The process was complemented with a top-down (deductive) approach in that we specifically organized the codes (findings) according to how participants reacted to each of the four features. This first phase resulted in a list of codes with associated quotes categorized under the four features, which were later re-grouped and organized into potential themes. This qualitative analysis complemented the quantitative results drawn from adherence, performance, and usage logs.

7.4 Findings

Throughout the four-week study, TandemTrack collected 428 exercise sessions (217 sit-up & 211 push-up sessions), 2,892 interaction logs with the mobile app, and 445 utterances with the skill. In addition, we collected 66 weekly diary entries, in which participants described their experience (including usability issues) with TandemTrack and reasons for missed exercise.

7.4.1 Exercise Behavior: Adherence and Performance

The two groups—M and MA—did not differ in their exercise completion (see Table 7.2), longest streak length [3] (M group: 10.90, MA group: 8.44), and average streak length (M group: 8.27, MA group: 6.02). Therefore, we combined the two groups for the remaining analyses on examining the factors affecting their exercise adherence and change in performance.

M group	**M-1**	**M-2**	**M-3**	**M-4**	**M-5**	**M-6**	**M-7**	**M-8**	**M-9**	**M-10**	**M-11**	**Avg.**
App	28	20	28	17	24	12	27	13	15	15	17	19.63
MA group	**MA-1**	**MA-2**	**MA-3**	**MA-4**	**MA-5**	**MA-6**	**MA-7**	**MA-8**	**MA-9**	**MA-10**	**MA-11**	**Avg.**
App	19	3	15	8	4	13	10	8	12	7	22	11 (57%)
Skill	0	17	2	5	20	0	14	11	6	15	1	8.27 (43%)
Total	19	20	17	13	24	13	24	19	18	22	23	19.27

Table 7.2: The number of exercise sessions each participant completed with the TandemTrack app & the skill.

As the study progressed, the performance for sit-ups has increased for 11 participants, decreased for two, and stayed the same for nine. Participants who completed more sit-up sessions were more likely to increase their sit-up performance (OR = .52, p = .04).

[3] The number of days that participants completed exercise sessions in a row.

On the other hand, the performance for push-ups has increased for eight participants, decreased for one, and stayed the same for 13. Participants who completed more push-up sessions did not necessarily improve their push-up performance. Regarding the change of sit-up/push-up performance over the study period, there was no difference between the two groups. Other factors, such as participants' longest streak length, average streak length, self-reported motivation level, and consistency in exercise time, did not affect the change of their sit-up/push-up performance.

Based on weekly diaries and interviews, we found five major reasons why participants missed exercise: busyness and prior commitments (n = 13); physical difficulties (n = 6); forgetfulness (n = 6); procrastination (n = 6); and missing a reminder (n = 4).

7.4.2 Usage of TandemTrack

Table 7.3 shows how participants in both groups engaged with different features of TandemTrack. In comparison to M participants (M = 26.36, SD = 6.82), MA participants (M = 20.72, SD = 6.59) opened the TandemTrack app less often, $t(20) = 2.20$, $p = .04$. This result is expected because MA participants could also use the skill, which provided the same features. In what follows, we focus on what MA participants preferred between the app and the skill regarding the four key features (i.e., exercise regimen and data capture, feedback, reminder) and reasoning behind their choice.

Engagement	M	MA	
	app	app	skill
# Opening TandemTrack	26.36	20.72	12.90
# Exercise sessions	19.63	11	8.27
Average time spent on reviewing feedback (seconds)	12.70	6.45	NA
# Swiping the daily feedback	37.09	31.72	NA
# Swiping the aggregated feedback	44.63	33.45	NA
# Tapping the mobile notification	13.09	7.81	NA
Average time elapse between reminder and exercise (minutes)	296	337	252
# Editing exercise data	0.09	1	NA

Table 7.3: Descriptive summary of participants' engagement with TandemTrack during the entire study period.

7.4.2.1 Exercise Regimen & Data Capture

In total, MA participants completed 212 exercise sessions, 57% of which were completed and captured using the TandemTrack app (see Table 7.2). We observed a large differences among the MA participants in what they used to exercise and capture data. Out of 11 MA participants, five used the TandemTrack skill more than the app for exercise regimen and data capture. These five MA participants had better exercise adherence (completed session: 21.8, longest streak length: 11, average streak length: 8.9) than the other six MA participants who used the TandemTrack app more (completed session: 17.2, longest streak length: 6, average streak length: 3.6), though we did not run a statistics test due to the small sample size. Notably, two MA participants (MA-1 and MA-6) did not use the skill at all, which we explain later.

During the interview, participants in both groups reported that the exercise regimen on the TandemTrack app easy to follow. They did not report any difficulty in data capture using the app. MA participants shared their rationale for choosing between the app and the skill to exercise during the interview: their decisions were influenced by their personal preference, proximity to the Echo Dot, their exercise environment, the social context around them, and the technical issues they encountered using the Echo Dot.

Personal Preferences: MA-2, MA-5, MA-7, and MA-10 showed a strong preference for the TandemTrack skill over the app for performing and capturing exercise, as they discovered the benefits that voice interaction offers. First, the hands-free interaction made the data capture more convenient. MA-10, who preferred using TandemTrack skill especially for push-ups, explained that "*When doing push-ups, you're putting a lot [of] strength to your hands, so it's a bit difficult to type [on the phone], but Alexa makes it easier*." In addition, MA-2 and MA-7 called out that the voice-based exercise regimen helped reduce distraction, so that they could focus on their performance. For example, MA-2 remarked, "*I don't have to look at my phone, but just listen to it [the Echo Dot] and do my exercise. It was less distracting, and I got to focus on my body and my performance.*"

On the contrary, MA-1 and MA-6 resisted interacting with the Echo Dot, because they did not feel comfortable talking to Alexa. MA-1 explained that "*This is not about privacy, I just did not feel comfortable talking to a personified machine. If it doesn't have name, maybe I'll try to talk to it.*" Similarly, MA-6 noted that "*It's something like almost human but not. The way to interact with it is not intuitive, it's awkward.*"

Proximity: Participants usually exercised in their home, but occasionally, they also reported exercising at the gym, friends' home, private rooms in the library, or in their offices. At home, some participants always exercised at the same place (e.g., bedroom) while others switched places (e.g., living room and basement) depending on their convenience. It was common for participants to exercise with the TandemTrack skill if they happened to be close to the Echo Dot when they received the exercise reminder. For example, MA-9 said "*I have my phone with me all the time, but if I was there when Echo reminded me, I would use it because it's convenient.*"

Exercise Environment: To effectively exercise with the TandemTrack skill, participants needed enough space to do sit-ups and push-ups while being proximate enough to the Echo Dot. For example, MA-10 rearranged her room by moving a desk next to her bed, and found it was no longer easy for her to exercise using the skill: "*This narrowed the space between the desk, where I put Alexa, and my bed. So I had to do exercise near the door of my room, but I couldn't use Alexa anymore—it's not close enough*."

Social Context: When the Echo Dot was placed at a location where other house members could access, participants worried that exercising with the TandemTrack skill might interrupt others. For example, MA-8 and MA-9 tended to exercise using the TandemTrack skill when other house members were not around, because they did not want to bother others by speaking to Alexa in the early morning or late night. Although most MA participants did not bring up privacy concerns while exercising with the TandemTrack skill, MA-4 emphasized that exercise is "*a personal, and private activity*," thus she did not want to exercise with the skill when other people were around: "*I'm a shy person, and I don't want to speak out loud to let others know that I'm doing this exercise*."

On the other hand, how other house members interacted with the Echo Dot also affected participants' tendency to use the skill. For example, MA-3 herself was not concerned about exercising using the skill, but the way her son interacted with Alexa annoyed her husband; as a result, she used the TandemTrack skill less often: "*ever since my four-*

year old son found he could ask Alexa 'Knock Knock Jokes,' he kept shouting at it every day. My husband got really annoyed so he unplugged the Echo a couple of times."

Speech Recognition Error: Six MA participants (3 international, 3 native English speakers) faced speech recognition error, which discouraged some of them from using the skill. Sometimes Alexa could not recognize the command "*Start TandemTrack*" accurately; in those cases, Alexa either responded with "*I'm not sure*" or started playing a random music. The latter was more frustrating than the former because participants had to first stop the music before trying to start the TandemTrack skill again. Speech recognition error also occurred when entering data. In such cases, participants had an option to use the mobile app to edit the data entries. In comparison to M participants (M = 0.09, SD = .30), MA participants (M = 1, SD = 1.34) edited their exercise data more frequently using the app: $t(20) = -2.19$, $p = .04$.

Because of these errors, some MA participants had an impression that the voice interaction was fragile—once an error occurs, there is no way to retract. For example, it took over 10 times for MA-11 to successfully invoke the TandemTrack skill during his first tryout. When the skill failed to recognize his repetitions accurately, he corrected the number in the TandemTrack app later, and decided to use the mobile app without trying the skill again: "*I think it just doesn't pick up my accent.*" Similarly, MA-10 wished the exercise reminder from Alexa could specify which exercise (i.e., sit-up or push-up) to perform before she invoked the TandemTrack skill, because "*I'm afraid of making mistakes on Alexa, so it would be better to start the skill and finish the exercise straightly instead of start the skill, get informed the exercise (type), get ready, and then start it again later.*"

7.4.2.2 Exercise Feedback

I found that participants who spent more time reviewing exercise feedback on the mobile app completed more exercise sessions ($b = .41$, $p = .04$), achieved a higher longest streak length ($b = .70$, $p = .02$) and a higher average streak length ($b = .07$, $p = .03$). While looking at how participants used TandemTrack's feedback features, I found that in comparison to M participants ($M = 12.70$, $SD = 6.41$), MA participants ($M = 6.54$, $SD = 4.09$) spent less time on the app reviewing exercise feedback in seconds, $t(20) = 2.73$, $p = .01$. Although MA participants had an option to check the exercise feedback through the skill's voice interaction in a Q&A manner, not many participants used it: over the course of 28 days, only three MA participants asked five questions in total.

Figure 7.3 shows *when* participants checked the feedback in relation to the exercise time. A majority of M participants and the four MA participants who heavily used the app for exercise (MA-1, MA-3, MA-6, MA-11) frequently reviewed their exercise feedback on the TandemTrack app around their exercise time, as M-1 noted, "*I just naturally looked at the chart after exercise*." Because they were already interacting with the TandemTrack app—either to initiate the exercise or to finish logging the exercise, checking the feedback was easy to tag on during this opportune moment. In addition, even after exercising with the skill, MA-7 and MA-10 would "*check whether the data was stored correctly*" on the mobile app, simply because "*opening the app on the mobile phone is easier than starting TandemTrack (skill)*."

On the other hand, exercising with the skill did not necessarily prompt MA participants to check the feedback neither from the app nor from the skill (with the exception

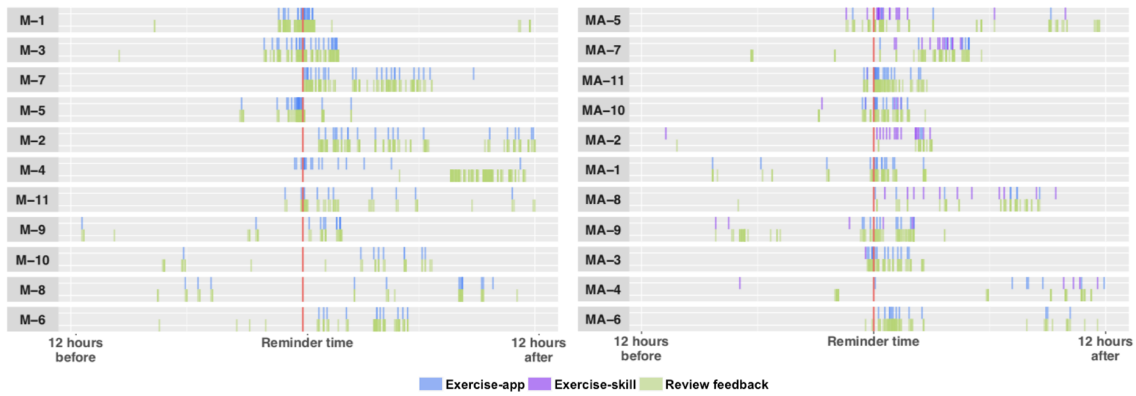


Figure 7.3: The time distribution of participants' interaction log with the TandemTrack app and their exercise entries (on the TandemTrack app & the skill) over the study period. Each row represents each participant's data, in an descending order of their complete exercise sessions.

of MA-5 who liked the feedback on the app, and MA-7 who wanted to check whether the data was stored correctly). This behavior was partly due to the speech-only, ephemeral nature of the interaction that does not provide visual cues, which was exemplified by MA-2: "*it* [checking the feedback] *just didn't pop up into my mind.*"

Participants in both groups acknowledged that the feedback on the TandemTrack app was informative to know how exercise progressed and encouraged them to stay motivated in the four-week challenge. For example, M-2 mentioned, "*I felt accountable when I saw this type of visualization. When I saw that I missed a day, I kind of want to keep doing it more so I don't miss anything in the future.*" But sometimes, the number of repetitions did not necessarily reflect participants' actual performance. For example, M-7 mentioned: "*to challenge myself, I intentionally did the sit-ups with a harder posture with my feet in the air. That's why the number doesn't look great.*" In addition, the performance-focused feedback (see C2 and C3 in Figure 7.2) sometimes made participants stressful, as M-4 noted that "*I don't like seeing the chart showing how you're doing, because I know that I was not that good.*"

MA participants rarely asked the TandemTrack skill questions about their exercise data. During the interview, MA participants shared various reasons for not using the voice interaction for feedback. Seven MA participants explained that they did not feel the need to ask Alexa questions given that the feedback on the app already had sufficient information. Two participants forgot what voice commands to use for asking questions (MA-3, MA-4), and one found the responses from Alexa not insightful (MA-10).

7.4.2.3 Reminder

I found that participants who exercised closer to their reminder time consistently completed more exercise sessions (b = -.01, p = .03), and achieved a higher average streak length (b = -.01, p = .04), but not necessarily a higher longest streak length. Participants in both groups found that the reminders on both devices were helpful, especially when the reminder time was a part of their daily routines. For example, M-1 used the exercise reminder as his wake-up alarm, which turned out to be effective: "*It (the reminder) is critical at the beginning, especially for the first week, but once I get into the habit—getting up at 8:00 AM every day, I don't really need that*." It is noteworthy that M-1 completed all of the exercise sessions (28 days). However, the initial setting of reminder time did not always work for everyone: participants often found themselves not ready to exercise when receiving the reminder. For those who received many notifications on their phone, the exercise reminder from the app was easy to be ignored.

Although there was no difference between the two groups regarding the consistency of their exercise time, I found that for MA participants, the exercise sessions completed on the TandemTrack skill (*M* = 252, *SD* = 229.76) were closer to their reminder time than the sessions completed on the mobile app (*M* = 337, *SD* = 325.29), $b = 208.17$ [4], $p = .005$. During the interview, MA participants acknowledged that the Alexa reminder sent from Echo device was more noticeable and accountable than the mobile notification. For example, MA-5 noted, "*it's easy to swipe the notification (on the mobile phone) unintentionally. The voice (reminder) is more accountable in making you be aware of, it's time to exercise*." Even MA-1 and MA-6, who resisted using the Echo Dot, acknowledged the effectiveness of the voice reminder: "*I do feel weird about it, but it did get me to exercise because I'm not always on my phone*" (MA-1). In addition, to get ready for exercise, MA-5 and MA-10 hoped that the Alexa reminder could specify which type of exercise (sit-ups or push-ups) should be performed, so that they do not need to look it up on the app.

7.5 Implications

In this section, I discuss the lessons learned from the study and implications for effectively combining a smart speaker and a mobile app.

7.5.1 Reflecting on the Exercise Adherence and Performance

The results showed no difference between M and MA group regarding their exercise adherence and performance, which could be due to two reasons. First, there were large individual differences in MA participants' choice between the skill and the app, affected by many factors described in 4.3.1 (e.g., personal preferences, proximity, social context). Four MA participants (MA-1, MA-3, MA-6, MA-11) heavily used TandemTrack's mobile

[4] *b* refers to regression coefficient.

app, and thus their usage was similar to that of M group. On the other hand, five MA participants (MA-2, MA-5, MA-7, MA-8, MA-10) used both the skill and the app, with a clear preference toward the skill. Given that the goal of this work is to examine ways in which we can complement the two modalities in designing technologies for exercise support, it is much more important to look at *how* participants interacted with each of the modalities and *why*, rather than simply looking at the adherence and performance. I believe this work contributes to such an understanding.

Second, participants' behaviors might have been affected by other factors that I have not captured in this study. As shown in Table 7.2, three M participants (M-1, M-3, M-7) achieved the adherence rate of 93% or higher whereas none of MA participants achieved such high adherence rates, which leads to the question that if I have neglected other important factors in assigning participants to groups (e.g., habit strength [285], self-regulation [286]), or whether the result is due to the small sample size. Those who stick to the 28-day exercise regimen are quite unique, and it would be interesting to examine the factors that can contribute in predicting such a behavior. Reflecting on the group assignment, although M group' app use data served as an important baseline, the findings around MA participants' app and skill use revealed more interesting insights. Therefore, in future studies that compare a novel system with a conventional one, assigning more participants to use the novel system can be an effective strategy to help researchers know more about the novel system.

Given the results, I learned that having people commit and stick to a simple exercise regimen—which takes 90 seconds per day, with the total interaction being less than 5 minutes—is very hard. Among the reasons why people missed the exercise (page 6), not many can be addressed by technologies, especially when people are reluctant to do the exercise. Simply providing a multimodal experience therefore would not motivate people to create a daily exercise routine. However, I believe that many interesting opportunities exist in *enriching* people's exercise experience by integrating multimodal interaction, if people are motivated to engage.

7.5.2 How Can Smart Speakers Complement the Mobile?

Here, I discuss the opportunities for smart speakers' voice interaction to complement the mobile app in delivering exercise regimen, facilitating data capture, enriching feedback, and increasing the effectiveness of reminders.

7.5.2.1 Optimizing Performance By Voice-Enabled Exercise Regimen

For fitness fanatics who are already familiar with the forms and routines, the visual interface can be distracting. I noticed that some of our participants paid a close attention to their forms and postures, and were keen to achieve high performance (M-2, M-7, MA-2, MA-7). While exercising using the TandemTrack skill, MA-2 and MA-7 reported that the audio-based exercise regimen helped reduce distraction, allowing them to focus on their performance. Therefore, a multimodal exercise regimen can be tailored by individuals' exercise knowledge and skill level [287]. For those who already know correct postures, speech interaction can serve as the primary communicate channel, to which the screen-based interaction can be added when it is necessary.

7.5.2.2 Facilitating Data Capture With Hands-Free Interaction

I found that hands-free speech interaction complemented the mobile app's capability in data capture, increasing the convenience for participants to record their exercise data while performing exercise, especially when they were intensively using their hands (i.e., doing push-ups). Although this study only focused on sit-ups and push-ups for the purpose of reducing the exercise complexity, speech interaction shows promising opportunities to support other exercises that require using hands (e.g., plank, body-pump, dumbbell).

On the other hand, speech interaction did not always work smoothly due to recognition errors; participants reported situations when Alexa could not accurately recognize their speech input, making them feel frustrating, and even lose confidence in using speech input. This finding echoes with prior research, which found that when a conversational agent fails to perform a task, people would lower their expectations on the agent's capabilities [288]. Thus, as we leveraged the smart speaker's advantages to complement the mobile app, it is equally important to leverage the strength of the mobile phone to complement the smart speaker. In this study, the option of editing one's exercise data using the mobile app helped participants correct their repetitions incorrectly recognized by Alexa, which suggested the benefits of using the mobile app to complement voice interaction's limitations. Going forward, when Alexa fails to recognize the voice command "Start TandemTrack," the mobile app should offer people an option to invoke the skill from their phone. Similarly, if the exercise session on the skill is interrupted, they should be able to resume the session with the mobile app.

7.5.2.3 Enriching Feedback Combining Voice and Visual

The results showed that the more time participants spent on reviewing their exercise feedback using the TandemTrack app, the better adherence they were likely to achieve. This finding corroborates previous research on using technologies to encourage physical activities, suggesting that feedback is important in promoting behavior change through increasing self-awareness and self-reflection [146, 289, 290].

When it comes to interacting with their exercise feedback, MA participants showed a strong preference for the mobile app over the skill. Although TandemTrack allows people to review their exercise feedback on the mobile app while listening to the voice feedback from the skill, We did not observe such use cases. In fact, some participants still chose to review their exercise feedback on the mobile app after completing exercise. We suspect that the low usage of voice feedback (from smart speakers) was due to the difficulties in discovering and remembering the voice commands as well as the lack of (1) coordination between the app and the skill in delivering feedback; (2) additional interesting information given the rich information provided by the mobile app; and (3) support for data exploration. To enrich individuals' experience with multimodal feedback, it is important to leverage the advantages of the two modalities in a synergistic manner for different scenarios. For data overview, we can use the mobile app as the primary interface, while enabling people to ask specific questions through speech input from the mobile phone [**?**]. In addition, instead of waiting for people to initiate a query, the skill can learn about individuals' interaction pattern with the app, and proactively prompt them with the type of data they are interested in, especially before or after exercise—"the critical reflection moments."

7.5.2.4 Delivering Effective Multimodal Reminders

Reminders are powerful nudges, as we found that participants who exercised consistently closer to their reminder time completed more exercise sessions and achieved better average streak length. Although the two groups did not differ regarding the consistency of their exercise time, We found that for MA participants, the exercise sessions completed on the TandemTrack skill were closer to their reminder time than the sessions completed on the mobile app. During the interviews, MA participants also acknowledged that in comparison to the mobile notification, the Alexa reminder is more noticeable and reliable. The biggest challenge in designing effective reminders is to reach people at the right time in the right place. Often times, participants ignored the mobile notification either because they accidentally swiped it or they had too many other notifications on their phone; for MA participants, they missed the Alexa reminder when they were not close to the Echo Dot. In the context of supporting consistent exercise, a "proximity-based reminders" might be helpful, which can determine the best modality to send the reminder based on people's proximity to the smart speaker and their mobile phone, within a certain range of the reminder time.

7.5.3 Technical Limitations of the Existing Platform

Several inherent limitations within the development platform (Alexa Skills Kit, or ASK) prevented me from designing an ideally "integrated" multimodal system. First, the current version of Alexa skills cannot send their own reminders; the only way to deliver notifications from third-party skills is through a visual cue (i.e., a spinning yellow light) on the Echo device, which made it hard to notice and to discern the source of the notifications. Therefore, we used Alexa's native reminder, which speaks out "time to exercise." In this way, participants were aware that the reminder was for TandemTrack, but we could not capture whether participants actually responded to the reminder because Alexa's reminder is one-way communication and is not tied with the TandemTrack skill. As a result, participants would still receive the reminder even if they had already completed exercise before the reminder time.

Second, Alexa's voice interface is vulnerable to speech recognition errors. For example, if the system fails to receive a valid speech input within the eight-second time-out window, the conversation session will automatically expire. In such a case, people need to initiate the session from the beginning to finish a task. In this study, exercising with the TandemTrack skill requires more than one round of speech interaction: participants needed to first invoke the skill, start training, and then report their repetitions three times. Within the interaction flow, any failure of speech recognition could impede a successful completion of the task.

Third, due to Amazon's limited support for third-party skills, we could not provide a truly integrated multimodal system, where people can use both the phone and smart speaker to control TandemTrack at the same time—for example, controlling the skill through the mobile app. Although some music apps (e.g., Spotify) allow people to control music streaming on the Echo device through both voice commands (e.g., "Pause") and the mobile app (e.g., "pause" button), this real-time synchronization is not well supported for other third-party skills.

For these limitations and yet having to rely on the ASK for the development of

TandemTrack, we could not leverage the synergy between the two modalities to their best. However, the lessons learned from the study are valuable in helping us understand what makes a technology an integrated multimodal system: one modality complements the other (and vice versa) when each of them lacking in a specific ability (e.g., editing exercise data on the mobile app). If the ASK allows, it would be better to show the captured data in real-time through the app when a person captures data through speech input, as a way to give assurance.

7.6 Chapter 7 Summary

In this chapter, I presented the design and evaluation of TandemTrack, a multimodal system comprised of a mobile app and an Alexa skill to support daily exercise. To examine how speech interaction complements and augments the mobile app, we conducted a four-week between-subjects study with 22 participants, followed by debriefing interviews. We identified the factors affecting participants' exercise adherence and performance change, and their preferences and use of TandemTrack's exercise regimen, data capture, feedback, and reminders with the two modalities. The findings uncovered the benefits and challenges of using speech interaction on smart speaker as a daily exercise assistant. The results generated interesting insights regarding how participants' personal preferences, proximity to the device, living environment, social context, and speech recognition errors influence their use of the TandemTrack app and the skill. With the lessons learned, I discussed design implications for how to combine the mobile app and speech interaction to build effective multimodal systems to encourage consistent exercise.

Chapter 8: Summary and Future Work

In this last chapter, I summarize prior chapters by highlighting the motivation, research approaches, contributions, and limitations. With the lessons learned, I discuss opportunities for future work and end this dissertation with a concluding remark.

8.1 Summary of Prior Chapters

8.1.1 Summary of Background and Motivation

In Chapter 1, I introduced the concept ***self-tracking***—"*a practice of noticing and recording the occurrence of one's target behavior*." In particular, I emphasized the benefits that manual tracking affords, such as increasing situated awareness and collecting rich personal contexts. By pointing out the limitations of touch input interfaces that are commonly used for manual tracking, I turned to the potential of speech input. I argued that the fast and flexible nature of speech input could complement touch input in supporting more efficient manual tracking. In this light, I presented five research questions under three themes: (1) *identifying design opportunities for multimodal self-tracking tools*; (2) *integrating speech and touch input on mobile phones to support self-tracking*; and (3) *examining the values of a smart speaker in supporting consistent self-tracking*.

In the first half of Chapter 2, I provided an overview of the main purposes of self-tracking: ***assessment*** purposes for healthcare providers to deliver treatment and ***therapeutic*** purposes for individuals to develop positive behavior change. I also summarized the benefits and limitations of different self-tracking approaches—automated, manual, and semi-automated tracking—that are equipped with various input modalities. Prior works shed lights on the dilemma wherein automated tracking limits individuals' awareness but most manual tracking approaches impose heavy data capture burden. In the meantime, the potential for speech input to lower the data capture burden is underexplored. Therefore, in the last half of Chapter 2, I reviewed previous research on speech-based data collection, including survey instruments and self-tracking systems. I also reviewed Natural Language Interfaces (NLIs) that process unstructured speech input into structured objects. I argued that self-tracking in real-world settings involves capturing different types of data, but many existing applications employ speech input to capture only a single type of data. Therefore, more research is needed to support multiple data capture by incorporating speech input.

In first half of Chapter 3, I described the three self-tracking contexts that I studied in this dissertation: food, productivity, and exercise. For food journaling, I described the diverse tracking needs of individuals and healthcare providers and highlighted the mismatch between these tracking needs and existing food tracker designs. I also described the importance of collecting various eating contexts to support a mindful and reflective food journaling experience and emphasized the heavy data capture burden that food journaling imposes on people. For productivity tracking, I reviewed prior works that aimed to boost productivity through tracking and displaying individuals' time spent on work. I also underlined the multifaceted nature of productivity and how contextualized productivity data can help researchers and organizations improve workplace wellbeing. For exercise tracking, I reviewed popular mobile fitness apps in the market and research prototypes aiming to promote physical activities. I also described how existing systems leveraged speech input to provide exercise guidance and pointed out that many of them do not support data capture. In the latter half of Chapter 3, I provided methodological foundations of this dissertation by describing how HCI researchers design and evaluate self-tracking technologies. For design methods, I focused on the rationales for co-designing with stakeholders and how they inspired my first step to explore opportunities for food tracker customization. For evaluation methods, I focused on how self-tracking technology is usually evaluated in three aspects: *examining behavior change*, *eliciting everyday interaction experience*, and *assessing data quality*. I argued that although behavior change is important, it is oftentimes challenging or even unfeasible to evaluate sustainable behavior change with limited resources. Just as importantly, understanding *how* and *why* a self-tracking technology influences people's daily life can make valuable contributions to the HCI and Personal Informatics fields.

8.1.2 Summary of Research Questions and Approaches

Centering around how multimodal data input (i.e., touch and speech) supports capturing personal data, I conducted four interconnected studies focusing on capturing different types of structured or unstructured data. I completed the studies in different self-tracking contexts, including food practice, productivity, and exercise (See Table 8.1). The

Research Question	Domain	Participants	Approaches
RQ1. What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems?	Food practice	Registered dietitians (N = 6)	**Co-design workshops** with dietitians using paper-based design widgets to create food trackers for patients with various dietary problems
RQ2. What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?	Food practice	People who are interested in learning about their food decisions (N = 11)	A one-week data collection study deploying **FoodScrap**, a speech-based food journaling app to capture everyday food practice
RQ3. How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes? **RQ4.** How does the input modality affect the data richness in unstructured input?	Productivity	Working graduate students who are curious about their time spent on school and work (N = 17)	Design, development, and deployment of **NoteWordy** (two weeks), a multimodal mobile app integrating touch and speech input to capture different types of productivity and break-related data
RQ5. How does a smart speaker complement and augment a mobile app in supporting consistent exercise?	Exercise	People who are motivated to do short strength exercise on daily basis (N = 22)	Design, development, and deployment of **TandemTrack** (four weeks), a multimodal system coupling a mobile app and an Alexa skill, in a between-subjects study to support in-home exercise training and tracking

Table 8.1: A summary of research questions, study domain, participants, and approaches.

following summarizes the research questions and approaches.

In Chapter 4, I described how I addressed **RQ1** (What are the design opportunities, from the perspective of healthcare providers, for multimodal data input to customize food trackers to support patients with various dietary problems?) through six individual co-design workshops with registered dietitians. The study took a qualitative approach, which acted as the first step to explore opportunities for designing multimodal self-tracking tools. The co-design workshops generated more than 30 unique tracking items, including food, reflection, activity, symptoms, and physical status. Depending on patients' dietary problems and dietitians' practice, the necessity and importance of these tracking items vary. Even for the same tracking items, they can be tracked with different timing and frequency, in different data formats and input modalities. While dietitians envisioned using most of the data for assessment purposes, they pointed out that some data (e.g., reflective thoughts) are captured for patients' own records, which were intended to help increase awareness and reflection instead of being sharing at clinics.

In Chapter 5, I created a speech-based food journaling app that collects people's daily food practice. To answer **RQ2** (What is the experience of capturing everyday food practice using speech input, regarding data richness and data capture burden?), I deployed FoodScrap in a one-week long data collection study (N = 11), during which participants recorded their food components, preparation methods, and food decisions through speech input. I also employed a post-study questionnaire to examine how participants perceived the data capture burden and conducted debriefing interviews to understand how speech input in FoodScrap affected their self-awareness and reflection. This work was mainly qualitative, focusing on eliciting participant's everyday interaction experience with FoodScrap and examining the richness of the collected data as a way to assess data quality. With speech input, participants detailed their meal ingredients and elaborated their food decisions by describing the eating moments, explaining their eating strategy, and assessing their food practice. They recognized FoodScrap facilitated situated reflection, but also expressed concerns about using speech input in public spaces or having to re-record the entries when errors occur.

In Chapter 6, I designed and developed NoteWordy, a mobile app integrating touch and speech input to capture different types of data. To answer **RQ3** (How do people use touch and speech input, individually or together, to capture different types of data for self-tracking purposes?) and **RQ4** (How does the input modality affect the data richness in unstructured data?), I deployed Notewordy in the context of productivity tracking with working graduate students for two weeks (N = 17), followed by debriefing interviews.

During the study, participants could choose between or combine touch and speech input to capture different types of data about their tasks and breaks. This work combines quantitative and qualitative approaches to examine participant's everyday interaction experience with NoteWordy, focusing on their modality preferences and data richness. Participants generated several input patterns with touch and speech input, which varied by the data type as well as their input habits, error tolerance for speech recognition issues, and social surroundings. Additionally, speech input was shown to reduce the time spent on completing the diary entries and enhance the data richness of free-form text.

In Chapter 7, I designed and developed TandemTrack, a multimodal system comprised of a mobile app and an Alexa skill to support daily exercise. To answer **RQ5** (How does a smart speaker complement and augment a mobile app in supporting consistent exercise?), I conducted a four-week between-subjects study ($N = 22$), followed by debriefing interviews. During the study, one group used both the mobile app and the Alexa skill to track their daily exercise and the other used the mobile app only. Combining both quantitative and qualitative approaches, this study focused on eliciting participant's everyday interaction experience with the TandemTrack app and the skill and the differences between the two groups regarding their exercise behavior. Although the results showed that the two groups did not differ in their exercise adherence and performance, I found that the TandemTrack skill largely enriched participants' exercise experience with its hands-free interaction. When it came to choosing between the app and skill to perform exercise and data capture, participants' decisions were influenced by their personal preferences, proximity to the device, the exercise environment, the social context around them, and the technical issues they encountered with smart speakers.

8.1.3 Summary of Contributions

This dissertation mainly contributes to the fields of Human-Computer Interaction (HCI) and Personal Informatics [1]. The contributions can be summarized to: (1) methodological contributions of structuring co-design workshops with healthcare providers to solicit design considerations for self-tracking tools; (2) design and implementation of two multimodal self-tracking systems—NoteWordy and TandemTrack—that integrate touch and speech input; and (3) empirical understandings such as customization dimensions of tracker design, and benefits and limitations of speech as an input modality to support self-tracking based on deployment studies, questionnaires, and interviews.

8.1.3.1 Methodological Contributions

Methodological contributions create new knowledge that informs researchers to carry out their work in the future [291]. In this dissertation, I made methodological contributions by structuring a new way of conducting co-design workshops with registered dietitians to generate design ideas for self-tracking tools targeting different patients. I propose the following methodological guidelines for structuring co-design workshops with healthcare providers, with the aims of best leveraging their expertise and collecting realistic design requirements:

- **Creating patient personas for contextualization:** Patient personas should include concrete information regarding age, gender, health symptoms, and goals to help

[1] In "Research Contributions in Human-Computer Interaction" [291], Wobbrock and Kientz described seven types of unique contributions that HCI researchers make. Each contribution type has key characteristics that imply how it is judged.

providers and researchers be contextualized in patients' experience.

- **Preparing flexible design widgets to facilitate the design process:** The design widgets should serve as building blocks to help providers get started quickly (e.g., predefined data fields) and modify their designs easily (e.g., paper-based widgets). It is also important to provide participants with opportunities to think beyond what researchers have prepared (e.g., blank widgets).

- **Considering healthcare providers' practicing style:** Involving providers from different backgrounds and specialized areas can help researchers understand how providers' practicing styles interplay with patients' characteristics to produce different designs.

- **Prioritizing design considerations:** While co-design workshops often open a wide range of possibilities without technical constraints, researchers should ask providers to prioritize their design considerations striking a balance between feasibility and desirability. This can be achieved by asking providers to articulate their design rationale and envision how they would use the system in their clinical workflow.

These methodological guidelines can assist HCI and Health Informatics researchers in structuring effective co-design workshops when working with healthcare providers.

8.1.3.2 Design and Implementation

Artifacts in HCI are inventions, including prototypes, systems, architectures, and toolkits, which facilitate new insights or compel researchers to consider new possibilities [291]. This dissertation contributes the following artifacts through design and implementation of multimodal self-tracking systems.

- **NoteWordy:** The design of NoteWordy's data capture interface illustrated how to effectively integrate touch and speech input on smartphones to capture different types of data for self-tracking purposes. Specifically, the ***local speech (LS)*** input allowed people to choose between speech and touch input for each data field, and the ***global speech (GS)*** input enabled people to capture multiple data fields collectively. NoteWordy's GS pipeline demonstrated how we can leverage existing NLP resources with additional heuristics to categorize natural language input into structured objects and how such pipeline can be applied to support self-tracking. Similar approaches can be used by other researchers to build and improve their own speech-based data capture systems. As one of the first systems that support multimodal self-tracking, NoteWordy also showed how speech input can reduce time spent on completing diary entries and enhance the richness of unstructured data.

- **TandemTrack:** To the best of my knowledge, TandemTrack is the first multimodal system that integrates a mobile app and a smart speaker to support the full cycle of exercise training and tracking with exercise regimen, data capture, feedback, and reminders. The exercise regimen and data capture mechanism within the TandemTrack skill demonstrated how the hands-free interaction facilitated in-home exercise training, which opened up opportunities for supporting more diverse and complicated workouts on smart speakers. The exercise feedback on the TandemTrack app, on the other hand, allowed people to explore and reflect on their data before or after exercise sessions, which highlighted the promise of proactively prompting people to review their exercise data during these "*critical reflection moments*."

8.1.3.3 Empirical Findings

Empirical contribution, which provides new knowledge through findings based on observation and data gathering, is the backbone of HCI research [291]. Through co-designing with registered dietitians and deploying three self-tracking systems in real-world settings, my dissertation makes the following empirical contributions:

- **Customization dimensions of tracker design:** Drawing from the co-design study with dietitians, I identified multiple dimensions to customize food trackers for different patients, including tracking items, data format, timing and frequency of tracking. By highlighting the dynamic nature of self-tracking in clinical settings, I provide design considerations for creating tracking templates that can be modified to meet the evolving tracking needs of patients and providers.

- **Benefits of touch input:** The main benefits of using touch input to capture personal data include: (1) being fast and easy for collecting structured data such as multiple choices and Likert scale; (2) mitigating privacy concerns in public settings; (3) helping people better organize their thoughts for capturing unstructured and complicated information; (4) easier to edit compared with speech input.

- **Limitations of touch input:** Despite the benefits, touch input is limited in supporting self-tracking due to: (1) heavy input burden for typing; (2) limited flexibility for capturing structured data, such as time (e.g., selecting the specific date, hour, and minute information) and multiple choices (e.g., selecting an item from a list without the opportunities to elaborate the response); (3) not being easy to use when people are doing hands-intensive activities (e.g., in-home exercise).

- **Benefits of speech input:** The main benefits that speech input offers include: (1) being generally perceived fast and convenient for capturing unstructured data (i.e, text); (2) enhancing the richness of unstructured data by encouraging people to elaborate their responses with specific details and additional contexts; (3) facilitating situated reflection when capturing activities that involve inner thoughts such as food decisions; and (4) saving time compared with the traditional touch input when it comes to entering the same amount of data fields.

- **Limitations of speech input:** Despite the benefits, speech input is limited in supporting self-tracking due to: (1) extra input effort to edit the data (e.g., re-recording the response) when mistakes were made; (2) mental load incurred by capturing long and complicated thoughts (e.g., food decisions, productivity rationale) if people do not have the information organized in their mind; (3) being constrained by social environments due to background noises, privacy concerns, or feeling embarrassed about disclosing personal information in public settings; (4) difficulty or discomfort in phrasing an utterance to include multiple data fields, especially if the sentence does not flow naturally; and (5) being vulnerable to speech recognition errors.

- **Additional values of smart speakers:** Although adding a smart speaker did not necessarily increase exercise adherence and performance, I found several values that smart speakers offer to enrich people's exercise experience: (1) helping optimize exercise performance by reducing visual distraction, especially for exercise fanatics; (2) facilitating data capture with hands-free interaction, especially for capturing hands-intensive exercise (e.g., push-up); (3) promoting consistency in exercise time, as the study showed that participants' exercise time tended to be closer to their reminder time when they chose to exercise with smart speakers.

- **Additional drawbacks of smart speakers:** The drawbacks of smart speakers in supporting in-home exercise training and tracking include: (1) limited ability to provide exercise feedback for people to revisit and explore their exercise data; (2) interaction experience can be affected by proximity to the device; and (3) people who are concerned about the personified feature of intelligent agents may resist using smart speakers.

- **The roles of feedback on speech input:** The four studies employed several ways to deliver feedback on speech input, and revealed the benefits and limitations of these feedback mechanisms: (1) audio recording lacking immediate visual feedback could encourage patients with dietary problems to frankly share their thoughts without feeling shame, especially those who experience eating disorders; (2) listening to one's own audio recordings on reflective thoughts (e.g., food decisions) could support people to reflect on past activities without the need to looking at the screen, although others might not like to hear their own voice; (3) real-time text transcriptions could be helpful in making people be aware of what is being recognized and support them to edit text entries via typing; (4) synthetic audio feedback sent by smart speakers provided little engagement, because people did not remember what questions they could ask and found that the aggregated visualizations provided by the mobile app were more helpful.

8.1.4 Summary of Limitations

This dissertation has several limitations, some of which are tied to a specific study, while others are pertinent to the small sample size and short evaluation period in the deployment studies. Here, I summarize these limitations, explain the tradeoffs behind my research design, and discuss what can be improved in the future.

In Chapter 4, the co-design workshops identified several dimensions of customizing food trackers from dietitians' perspective, but overlooked the perspective of patients. While the patient personas helped dietitians articulate how their designs address the tracking needs regarding patient's characteristics, it was inevitable that the personas reflected the perspective of dietitians more so than that of patients. Given that the focus of this work was to understand how to customize food trackers to meet healthcare providers' assessment goals, I prioritized dietitians' information needs rather than patients' needs. Looking backward, this study was a good first step that laid out opportunities for multimodal input to support self-tracking.

In Chapter 5, the data collection study showed how FoodScrap encouraged participants to detail their meal components and elaborate food decisions, as well as how it facilitated situated reflection. We suspect that the observed outcomes were due to both the expressive nature of speech input and the guided promoting questions (e.g., "*Why did you eat at this time rather than earlier or later?*"). However, it was unclear how much each of these two factors contributed to the results. Considering that the goal of this work was to understand the experience of capturing everyday food practice using speech input, I put people in a situation where they were prompted to use speech input to capture different aspects of food practice. Drawing from the qualitative findings, I showed how we can create a reflective self-tracking experience to collect rich data by employing speech input. To quantitatively examine the roles that speech input plays in the data capture process, it warrants factorial experiments to compare speech input with other input modalities (e.g., text typing) and carefully design study conditions with different prompting questions.

In Chapter 6, I used existing NLP resources along with a set of rules and keywords to capture multiple data fields from NoteWordy's global speech (GS) input, which cannot be generalized to data capture regimens other than the Productivity Diary and Break Diary used in the study. However, the two diaries in the study equipped with touch, LS, and GS input can be considered as a test bed to situate people to gather real-world experience in capturing personal data with touch and speech input. In the meantime, the pipeline that handles natural language input from GS can be extended by other researchers with domain-specific rules and keywords to build their own data capture systems.

In Chapter 7, TandemTrack relied on self-reported data to capture sit-up and push-up repetitions, which might not be accurate (e.g., participants could have entered inaccurate repetitions). Besides, due to the limitations within the development platform of Alexa skill, TandemTrack was not the most ideal multimodal system (See Chapter 7, Section 7.5.3). For example, the exercise reminder from the smart speaker was Alexa's native reminder, which was not tied to the TandemTrack skill. Therefore, the study could not track whether participants actually responded to the reminders or not. However, in exploring opportunities for supporting consistent exercise training and tracking with multimodal interaction, TandemTrack allowed people to interact with a smart speaker and a mobile app to perform daily exercise; from these interactions, we identified rich insights regarding how speech input on smart speakers influenced people's exercise experience.

Across the three deployment studies with FoodScrap, NoteWordy, and TandemTrack in Chapters 5–7, there are two additional limitations. First, I deployed the systems to only a small number of participants (ranging from 11 to 22), which may limit the generalizability of the results. In the FoodScrap study, the participants were female dominated (82%), thus a study with a larger population involving more male participants would likely generate more diverse results (e.g., identifying new food details or elaboration types). Similarly, in the TandemTrack study, each study group included only 11 participants, and most of them were college students (86%). Therefore, a larger sample size may allow us to observe other factors influencing people's exercise experience and even significant differences between the two groups. In the NoteWordy study, the statistic power of regression analysis may be limited by the sample size (N = 17). However, as I noted in Chapter 6, the number of observations in the regression models were based on how many diary entries participants produced. The power analysis showed that all the regression models reached over 90% power with a medium effect size, suggesting that the quantitative findings were valid. Although the R^2 of the regression models seemed to be low (.1 to .15), prior research has pointed out that due to the complex nature of human behaviors, a low R^2 associated with behavioral metrics still helps explain the variances of the dependent variable [292].

Second, all three studies lasted only one to four weeks, which made it difficult to eliminate the novelty effects and estimate the systems' long-term impact. In particular, I deployed FoodScrap and NoteWordy as data collection tools, and did not examine whether the two systems affected participants' eating behaviors or productivity level. However, the primary goal of my dissertation centers on examining *how* and *why* people perceive and use the self-tracking systems, along with multimodal data input, in particular ways. As exploratory research, the FoodScrap and NoteWordy studies focused on eliciting people's modality preferences, data capture burden, and data richness by interacting with systems on daily basis (See Chapter 3, Section 3.4.2), which made a set of artifact and empirical contributions to the HCI and Personal Informatics fields. In the future, the lessons learned from these studies can be incorporated to build a self-tracking system aiming to improve positive behaviors, and we can conduct long-term studies to examine how the system affects people's behavior and life. Additionally, instead of focusing on people who are first-time users of multimodal technology, researchers can conduct more formative studies by involving those who have been using or have used multimodal input to capture their personal data (e.g., voice journaling) for a longer time (e.g., more than 6 months), to understand how their experience evolves over time.

8.2 Future Research Agenda

The findings of this dissertation revealed several opportunities to continue supporting multimodal self-tracking as well as new avenues for future research. Here, I list the specific research topics that I plan to pursue.

8.2.1 Enriching The Full Cycle of Personal Informatics Through Multimodal Interaction

According to the stage-based model of personal informatics proposed by Li and colleagues in 2010, self-tracking systems are composed of five stages, including preparation, data collection, integration, reflection, and action [4]. My dissertation mainly focused on the data collection stage by examining how touch and speech work together as input modalities. However, little is known about how multimodal interaction can be applied to facilitate other stages of personal informatics, such as enabling reflection with feedback and supporting preparation by sending effective reminders. Inspired by participants' reaction during the studies, I see the following opportunities:

- **Tracking multiple target activities with speech input:** Drawing from the general positive experience with NoteWordy's GS input, we can extend the data capture mechanism to capture multiple target activities (e.g., exercise, food, mood) on smartphones. To further lower the input burden, people can enter their data with speech input from the lock screen without opening a specific app or diary, and then the system can automatically categorize the key values into corresponding entries (e.g., extract the exercise and food type and place the values to corresponding diaries from "*I ran 5 miles this morning and had some bread*").

- **Enabling text editing with touch and speech input:** In the NoteWordy study, people can easily edit their data using touch and local speech (LS) input (e.g., updating the selected item in multiple choices with touch input, updating the duration with LS by saying "*from 8 to 9 p.m.*"), but it can be challenging to edit the details in unstructured text fields. Oftentimes, people need to either (1) precisely move the cursor within the target area, remove the text, and type in updated text or (2) re-record the entire information even though only a small part of it needs to be updated. To make text editing easier, we can combine touch and speech input by enabling people to first select the target text area and then record only the updated information via speech input (e.g., select "*for the master*" and say "*for the semester*"). We can also explore command-based approaches by allowing people to use speech input to specify the portion that needs to be corrected along with the updated information (e.g., say "*replace 'master' with 'semester'*") [293].

- **Investigating the roles of visual and auditory feedback on speech input:** The four studies revealed contradictory perspectives of how feedback on speech input affected people's self-tracking experience. For example, the co-design study with dietitians and the deployment of FoodScrap showed that audio recording's lack of visual feedback could encourage people to frankly share their thoughts and enable situated reflection. However, lack of visual feedback could make it difficult for people to keep their train of thoughts and edit their data. In the NoteWordy and TandemTrack studies, visual feedback (e.g., real-time transcription and aggregated exercise data) was recognized to be helpful. We suspect that the contradiction was due to a mix of reasons, such as the sensitivity of the data (e.g., not wanting to directly see past negative behaviors), participants' trust of the system's speech recognition accuracy (e.g., feeling assured with text transcripts), and how their self-reflection occurred (e.g., through retrospectively exploration *vs.* thinking aloud in-situ). In addition, FoodScrap and NoteWordy provided only one type of feedback, which might have limited participants' desires to engage with their data in other ways. Therefore, future research can investigate multiple ways of combining visual (e.g., raw transcripts, extracted key words, visual charts) and auditory elements (e.g., original recordings, pitch, tone, background sounds) to support individuals to revisit, explore, and reflect on their personal data in various contexts.

- **Structuring feedback from natural language input based on different information needs:** Although speech input can help people collect rich contextual information related to the target activity, these contexts were identified and categorized by researchers based on the research goals, and therefore are (1) not scalable when the amount of data increases and (2) not applicable when the assessment needs change. While existing NLP techniques can extract domain-specific information with corresponding training data or rule-based heuristics, the information needs of individuals, healthcare providers, and other stakeholders (e.g., researchers) can be important parameters to include. In food journaling practice, for example, healthcare providers may need to assess individuals' nutrient intake, individual themselves may want to know whether they made progress toward their dietary goals, and food science researchers may be interested in how individuals make their food choices. In the FoodScrap study, these key values can be extracted from the same source of natural language input, suggesting that future systems can structure and present the data in different ways depending on the information needs of different groups.

- **Supporting data exploration with touch and speech interaction:** In Chapter 7, the TandemTrack Alexa skill allowed participants to ask basic questions about their exercise through a Q&A manner. However, participants rarely used this feature because (1) they were unsure about what questions to ask; and (2) they could quickly forget the details of the data due to the ephemeral nature of speech output. Instead of relying on the speech-only communication, visual aid can be incorporated into the process of delivering feedback. For example, smart speakers with a screen (e.g., Echo Show [294]) can provide visual charts showing the overview of one's data, from which people can ask more detailed questions via speech input (e.g., general trend, best performance).

- **Designing multimodal reminders across devices:** Findings from the TandemTrack study showed that participants' exercise time tended to be closer to their reminder time if they exercised with the smart speaker. While acknowledging that reminders sent from the smart speakers were more noticeable than mobile notifications, participants could miss these reminders if they were not close to the smart speaker. To design effective reminders to reach people at the right time in the right place, multimodal systems in the future can take people's proximity to each device into account so as to determine the most appropriate modality to send a reminder. In addition to mobile notification and smart speaker reminder, researchers can consider other modalities such as ambient display or shape changing interfaces [295].

8.2.2 Examining The Evolving Tracking Needs in Clinical Settings

Chapter 4 highlighted that self-tracking in clinical settings is a dynamic process: healthcare providers' and patients' tracking needs could change as the treatment progresses, suggesting that tracking tools should support flexible modification to the data capture regimens. Although this can be achieved by creating tracking templates that allow providers and patients to customize their tracking items and data format, other issues could arise: (1) how to ensure that both parties are on the same page when the tracking needs of one party change; and (2) how to provide aggregated feedback based on previous data, once the tracking regimen is updated. Without addressing these questions, it can be difficult to design tracking templates that can be used by providers and patients. Therefore, we first need to understand how tracking needs evolve in clinical settings, including what triggers the changes, what needs to be updated (e.g., tracking item, data format, input modality, data sharing preferences), and how providers communicate with their patients regarding the changes. An important next step is to understand both providers' and patients' perspectives on how they manage their evolving tracking needs and historical data, as well as associated challenges. This can be done through formative studies such as focus groups, field observations, or diary studies.

8.2.3 Investigating The Roles of Personified Agent in Self-Tracking

Conversational agents (CAs) have been gaining popularity in recent years. Particularly among health and wellbeing applications, CAs have been used as self-care systems through actively initiating health-related conversations with people and increasing their self-compassion [155]. Interestingly, Chapter 7 revealed diverging attitudes toward the TandemTrack skill as a CA: some participants felt encouraged to exercise with the skill, but others resisted using it because they did not feel comfortable interacting with a "*personified device*." Although we attributed this finding to personal preferences, it warrants more research to investigate what influences such preferences and what kinds of personified features we should include or exclude in CAs to support self-tracking. Referring to the theories of anthropomorphism (e.g., friendliness, need to disclosure, personality) [296, 297], future research may aim at understanding how people use the personified features on CAs for different purposes (e.g., asking for driving routes *vs.* tracking personal data) and then examine how these features can facilitate or impede self-tracking activities.

8.2.4 Multimodal Data Capture Beyond Touch and Speech Input

Although Chapters 5–7 showed that speech input can facilitate easy, fast, and reflective data capture, it is not always desirable to use speech input for capturing personal data due to environmental, social, and privacy concerns. To help people capture their data in a more intuitive way, future research can consider other input modalities beyond touch and speech input. For example, finger gestures afford intuitive expression of simple semantics, such as conveying "*positive*" messages with thumbs up and "*negative*" messages with thumbs down. Gestures are less invasive compared with speech input, and can be embedded in mobile phones, computers, and other smart home devices. In the context of mood tracking, when one feels not comfortable using speech input, using gesture input instead may help capture positive or negative mood. Such a multimodal self-tracking system composed of speech, touch, and gesture input across multiple devices will provide people more freedom to choose their preferred modality at different times depending on the social context and the type of data they intend to capture.

8.2.5 Reaching The Disability Community

In previous chapters, I deployed the self-tracking systems to individuals without disabilities as a first step to examine the feasibility of capturing personal data through multimodal input. Nevertheless, a multimodal system equipped with speech and touch input holds great promise for disabled individuals (e.g., people with visual or motor impairments). Some people in these groups also have the needs to monitor their health by tracking a variety of data. For example, research showed that people with low vision have been managing their eye health using mobile apps for vision test and eye exercise, which involves collecting photos and self-reported Likert scales [298]. However, existing technologies do not fully match their interaction capabilities (e.g., providing too much text information, requiring precise scale selection). As such, more research is needed to examine disabled individuals' tracking needs and how multimodal input can help them capture different types of data through formative studies (e.g., co-designing with individuals and their caregivers) and technology deployment in their natural living environment.

8.3 Concluding Remarks

Motivated by the importance of manually tracking personal data and the challenges it poses, I sought to support efficient manual tracking by integrating speech and the traditional touch input in this dissertation. Through designing, building, and evaluating different ways of incorporating speech input into self-tracking systems, the most important lesson I learned is that people's preferences for input modalities are affected by a mix of internal (e.g., previous input habits, error tolerance) and external (e.g., data type, social surroundings) factors. Speech as a fast and natural input modality can help individuals collect rich data faster, but may raise privacy concerns or extra mental load that can be mitigated by touch input. The key takeaways echoed the thesis statement: multimodal systems equipped with touch and speech input allow people to take the advantage of both modalities, which greatly increases the flexibility of data input.

Going forward, as speech recognition, NLP, and other multimodal interaction technologies continue to advance, I envision that future self-tracking systems will better leverage people's natural interaction capabilities to engage them with personal data tracking. I hope this work can inform and inspire other researchers working in the growing body of personal informatics to design multimodal systems that support rich, low-burden, and reflective self-tracking experiences.

Appendix A: Study Material for Co-Designing with Dietitians

A.1 Scripts For The Co-Design Session

A.1.1 Interviews Questions for Pre-Design Activity

Q1. Can you describe the patients you usually see, and a typical workflow of their visit?

Q2. Can you explain how you provide care to your patients?

Q3. How do you employ food diaries to your patients?

Q4. How do you usually motivate your patients to track their data?

Q5. For different types of patients, what are the important data for them to track?

Q6. In this study, we are going to design food trackers for your patients together. Before we move forward, can you describe two types of patients that you commonly see, and give us some information about their age, gender, eating goals, and health conditions?

A.1.2 Design Session

Now let's pick one patient persona *X*, and use our paper widgets to create a food tracker for this patient. During this process, think about questions such as:

Q1. What do patients *X* need to track about their food?

- food items (focus on only certain food items, or everything they eat?)

- serving size
- preparation method
- nutrition component (e.g., protein, fat, sugar)
- time of eating

Q2. Among these items, what is the most important thing? Why?

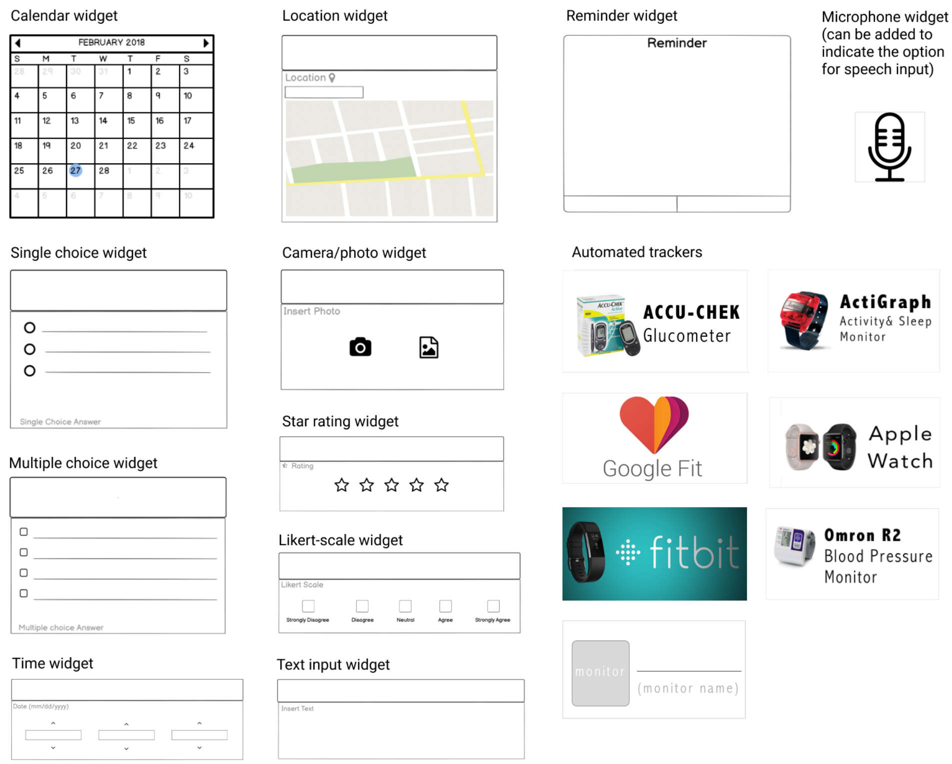
Q3. Anything they need to track together with their food intake?

- blood pressure
- weight
- sleep
- mood
- medication
- health symptoms

Q4. Anything they need to track together with their food intake?

Q5. How often do they need to capture this information?

A.2 Design Widgets



A.3 Debriefing Interview Questions

Q1. Let's reflect on the trackers we just designed. Can you summarize the some of the key design rationale of designing food trackers for different patients?

- What are the important factors to consider when designing trackers for different patients?
- What are the challenges for designing trackers for different patients?
- If we can develop a technology like that, what would you recommend to other experts in designing such trackers for their patients?

Q2. Can you talk about the experience of designing food trackers with the paper widgets?

Q3. Is there anything that the paper widget couldn't provide, or was not enough for [PATIENT TYPE] to track their food?

Q4. How are you going to use these data collected? What feedback would you like to provide to your patients?

Q5. Would this be good if you can receive these data from your patients and communicate with them in real time? Why or why not?

Q6. In your work practice, what is a rough percentage of patients who adhere to their treatment plans?

Q7. How do you deal with the adherence issues?

Q8. Would you want this app to be the primary form of tracking? Why or why not?

A.4 IRB Approval letter

UNIVERSITY OF MARYLAND
INSTITUTIONAL REVIEW BOARD

1204 Marie Mount Hall
College Park, MD 20742-5125
TEL 301.405.4212
FAX 301.314.1475
irb@umd.edu
www.umresearch.umd.edu/IRB

DATE:	February 2, 2018
TO:	Eun Kyoung Choe
FROM:	University of Maryland College Park (UMCP) IRB
PROJECT TITLE:	[1132164-2] Supporting personal data tracking and data exploration
REFERENCE #:	
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	February 2, 2018
EXPIRATION DATE:	November 7, 2018
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *7*

Thank you for your submission of Amendment/Modification materials for this project. The University of Maryland College Park (UMCP) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Prior to submission to the IRB Office, this project received scientific review from the departmental IRB Liaison.

This submission has received Expedited Review based on the applicable federal regulations.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of November 7, 2018.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Unless a consent waiver or alteration has been approved, Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

Appendix B: FoodScrap Study Material

B.1 Study Tutorial

Capturing Everyday Food Decisions

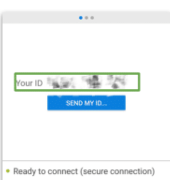
Pre-Study Tutorial

Agenda

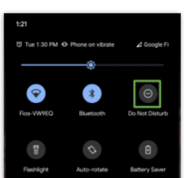
- Setting up phone screen sharing
- Study goals & procedures
- Setting up the study app
- Study tasks & Practice
- Important notes
- Questions

2

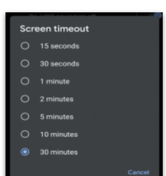
Setting up phone screen sharing



Launch TeamViewer QuickSupport and share the 10-digit code



Make sure the notification is off by turning on the "Do not disturb" mode



Fully charge your phone and set your screen timeout to maximal (Settings > Display > Screen timeout)

3

Study Goals

This is a **data collection** study aiming at

- Collecting how people make daily food decisions regarding what, when, and how much to eat
- Exploring how technologies can facilitate data collection on everyday food decisions

We are NOT going to judge or evaluate your food decisions

4

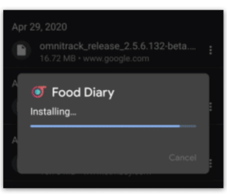
Study Procedures

Pre-study tutorial (45 minutes) → Data collection (7 days) → Post-study questionnaire (5 minutes) → Debriefing interviews (30 minutes)

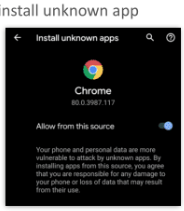
Study Setup

Installing the study app

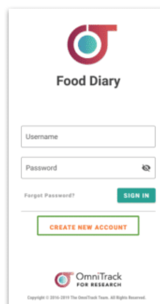
- Download the installation file using the link in your email
- Tap the downloaded file to install the study app

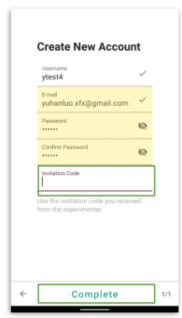


- If you encounter an issue while installing the app, try giving your browser the permission to install unknown app



Creating a study account



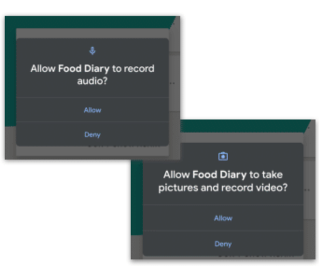


- Open the app and tap "Create New Account"
- Put your username, email, password
- Invite code: **2171**
- Tap "Complete" at the bottom when you're done

Grant permissions

To make sure that the study app runs properly, you need to grant the following permissions:

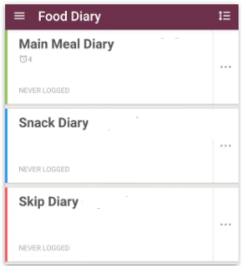
- Photo taking
- Audio recording
- Turn off the battery optimization (allow the app to run in the background)



Study Tasks

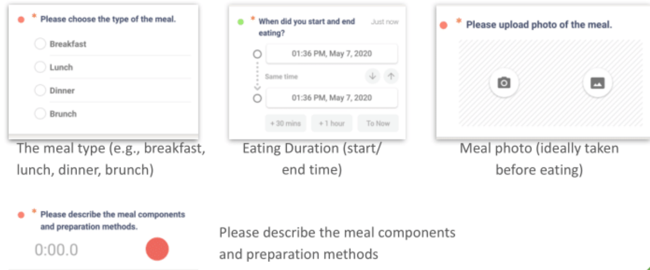
Three diaries

- **Main Meal Diary**
  - Capture breakfast, lunch, dinner
- **Snack Diary**
  - Capture anything you eat/drink other than your main meals
- **Skip Diary**
  - Capture any main meals you skip



Main Meal Diary

Food information



The meal type (e.g., breakfast, lunch, dinner, brunch)

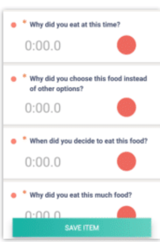
Eating Duration (start/ end time)

Meal photo (ideally taken before eating)

Please describe the meal components and preparation methods

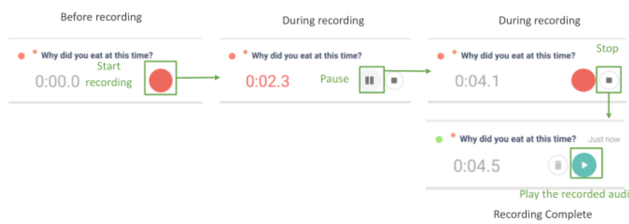
Main Meal Diary (Cont.) - Ideally captured right after eating

Food decisions: how you make food decisions regarding the specific meal, and how your the environment and people around you influence your decisions

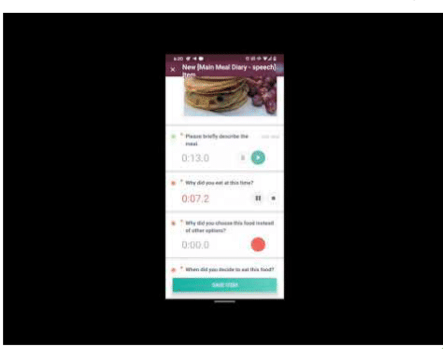


- Why did you eat at this time?
  - Why didn't you eat earlier or later?
- Why did you choose this food instead of other options?
  - Among the food that are available to you, why did you choose this food?
- When did you decide to eat this food?
  - Did you plan this meal earlier or just before eating?
- Why did you decide to eat this much food?
  - Why didn't you eat more or less?

Capture your food information using speech



Demonstration: Log your meal



Practice Session

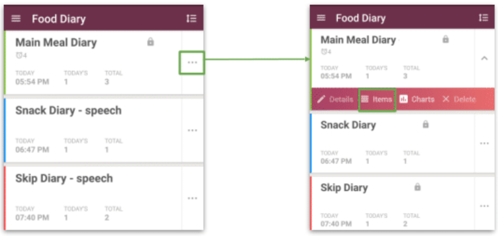
Think about the latest main meal you had and log it in the study app. While capturing your food decisions, refer the following information:

- Why did you eat at this time?
  - Why didn't you eat earlier or later?
- Why did you choose this food instead of other options?
  - Among the food that are available to you, why did you choose this food?
- When did you decide to eat this food?
  - Did you plan this meal earlier or just before eating?
- Why did you decide to eat this much food?
  - Why didn't you eat more or less?

Review your records

Tap the “three dots” icon on the diary you want to check

Tap the “items” to review your records



Snack Diary

What is considered as snack?
Anything you eat/drink other than the main meals (water is not included)

Snack

- Eating duration (start/end time)
- A photo of snack (can be uploaded before, during, after eating)
- Snack components and preparation methods

Snack decisions

- Why did you eat at this time?
- Why did you choose this snack instead of other options?
- When did you decide to eat this snack?
- Why did you decide to eat this much snack?

Skip Diary

Skipped meal

- The type of main meal you skip (e.g., breakfast, lunch, dinner)
- Note: brunch and snacks are not included

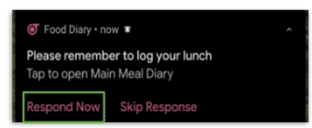
Skipping decisions

- When did you decide to skip this meal?
  - Did you plan to skip this meal earlier or you just made the decision to skip?
- Why did you skip this meal?
  - Why didn’t you have this meal as usual?

Reminder Setup

What are the regular times when you usually have breakfast, lunch, and dinner?

- One hour after your meal time
  - Breakfast reminder
  - Lunch reminder
  - Dinner reminder
- One hour after dinner reminder:
  - Summary reminder for any missing main meals



- You will not get a meal reminder if you have logged the meal before reminder time
- You will not get a summary reminder only if you have logged all the three main meals no later than one hour after your dinner reminder time

Important Notes

Study Expectation

- **What is considered as a valid entry?**
  - A meal within the same day (consumed or skipped)
  - An entry with **all** the fields filled (including the meal photo)
- **Snack Diary**
  - We encourage you to capture any snacks/drink you have
  - If you do not have snacks, you do not need to log the Snack Diary
- **How much detail to include in an entry?**
  - Try to capture all the food components in one meal and explain your reasons for choosing each food
  - If your food decisions are based on your daily habit, briefly explain how you developed this habit

Compensation

- **Your compensation is tied to your data entries**
  - $3 per main meal (consumed or skipped)
  - $7 bonus if all the three main meals are logged every day throughout 7 days (21 meals) --- maximal compensation: $70

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Breakfast	X	X	X	X	X		
Lunch	X	X	X	X	X	X	X
Dinner	X	X	X	X	X	X	X
Total	$9	$9	$9	$9	$9	$6	$6

Total compensation: $ 57 = 9 x 5 + 6 x 2

Dropout rules

- **Minimal requirement of data entry**
  - All three main meals for *at least five days* (consumed or skipped)
  - At least one main meal for all seven days (consumed or skipped)
- **You will be removed from the study if you do not meet the minimal requirement**

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Breakfast		X	X	X	X		
Lunch	X	X	X	X	X	X	X
Dinner	X	X	X	X	X	X	X

- **You will not get the study compensation if you are dropped out from the study**

Next steps

- The data collection starts **tomorrow** (from 7/7- 7/13/2020)
- You are encouraged to log all the meals/snacks you take at your best
- We will send you the post-study questionnaire by the end of the study, which is 7/14/2020
- We will schedule the debriefing interview with you after you complete the post-study questionnaire
- We will use TeamViewer QuickSupport again during the debriefing interview
- You will receive the study compensation at the end of the interview

Questions?

B.2 Post-Study Survey (Sub-Items From The User Burden Scale)

Please choose the option that best describes your experience in capturing your food components, preparation methods, and food decisions using speech input with the study app over the past week.

Q1. I need assistance from another person to capture my food components, preparation methods, and food decisions using speech input.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q2. Capturing my food components, preparation methods, and food decisions using speech input demands too much mental effort.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q3. It takes too long for me to do what I want to do while capturing my food components, preparation methods, and food decisions using speech input.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q4. It was difficult to learn how to capture my food components, preparation methods, and food decisions using speech input.

- Not at all
- A little bit
- Somewhat
- Very much
- Extremely

Q5. I spend too much time capturing my food components, preparation methods, and food decisions using speech input.

- Not at all
- A little bit

- Somewhat
- Very much
- Extremely

Q6. I use the study app to capture my food components, preparation methods, and food decisions more often than I should.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q7. Capturing my food components, preparation methods, and food decisions using speech input distracts me from social situations.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q8. Using speech input to capture my food components, preparation methods, and food decisions has a negative effect on my social life.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q9. Capturing my food components, preparation methods, and food decisions using speech input requires me to remember too much information.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q10. When I am capturing my food components, preparation methods, and food decisions, the study app presents too much information at once.

- Never
- A little bit of time

- Sometimes
- Very often
- All of the time

Q11. Using speech input to capture my food components, preparation methods, and food decisions makes me feel like a bad person.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q12. I feel guilty when I use speech input to capture my food components, preparation methods, and food decisions.

- Never
- A little bit of time
- Sometimes
- Very often
- All of the time

Q13. I am worried about what information gets shared when I use speech input to capture my food components, preparation methods, and food decisions.

- Not at all
- A little bit
- Somewhat
- Very much
- Extremely

Q14. The study app's policies about privacy are not trustworthy.

- Not at all
- A little bit
- Somewhat
- Very much
- Extremely

Q15. Capturing my food components, preparation methods, and food decisions using speech input requires me to do a lot to maintain my privacy within it.

- Never
- A little bit of time
- Sometimes

- Very often
- All of the time

B.3 Interview Questions

Q1. How was your experience of capturing your food and food decisions over the past week?

- In general, how do you describe your diet over the last week?
- Were there any health issues that you were concerned about while making food decisions?

Q2. When it comes to what to eat, what are the important considerations for you? Why do you think they are important?

- Were you aware of these decisions before the study (illustrate some food decisions that the participant has captured during the study)?
- Was this your typical diet? Was there anything special during this time (the COVID-19 pandemic)?

Q3. During the week of the study, what is the policy of your state?

- Are you free to dine outside?
- How comfortable are you in going to a restaurant/ordering delivery food?

Q4. Can you open your log history of Main Meal Diary, so we can look at it together while talking about your experience.

- In your food diary, it seems that you often [planned your meals beforehand/made the food decisions right before eating]; why was that?
- When did you usually capture your food and food decisions (e.g., did you log right after eating every meal, or did you log the meals after a while)?
- It looks like you eat fast/slowly most of the time, why is that?
- Can you share an example of a meal that is very difficult to describe?
- How do you estimate your eating duration?

Q5. (If many snacks were captured) I noticed that you captured many snacks during the study. When making decisions about having snacks, what are the differences from making decisions about your main meals?

Q6. (If little or no snacks were captured) I noticed that you did not capture many snacks during the study. Was that because you didn't have snacks, or you forgot about it?

Q7. In particular, how was your experience of using the speech input to capture your food decisions?

- In your post-study questionnaire, you noted that sometimes you felt [items from the User Burden Scale], can you tell me more about that?
- Anything you think can be improved?

Q8. How often did you listen to the audio that you have recorded in the study app?

- Why did/didn't you want to listen to the recorded audio?

Q9. Have you used the pause button while recording your decisions?

Q10. Had you ever used any type of food tracking apps or paper-based journal before this study?

- How was your experience with those apps?
- How was that experience different from/similar to this study?

Q11. (If eating with others is applicable) You mentioned in your earlier questionnaire that you live with your partner/family members. Overall, how do they/does she/he influence your food decisions?

- How often do you eat together?
- How actively were they in contributing to these food decisions?
- Were there any cases when you disagreed on what to eat? How did you usually resolve the disagreement?

Q12. What did you learn from the study about your food practice?

- What kind of feedback/information would you like to receive based on what you logged in your food diaries?

Q13. What did you learn from the study?

Q13. Do you have any questions for us?

B.4 IRB Approval letter

UNIVERSITY OF MARYLAND
INSTITUTIONAL REVIEW BOARD

1204 Marie Mount Hall
College Park, MD 20742-5125
TEL 301.405.4212
FAX 301.314.1475
irb@umd.edu
www.umresearch.umd.edu/IRB

DATE:	May 21, 2020
TO:	Eun Kyoung Choe
FROM:	University of Maryland College Park (UMCP) IRB
PROJECT TITLE:	[1132164-14] Supporting personal data tracking and data exploration
REFERENCE #:	
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	May 21, 2020
EXPIRATION DATE:	November 7, 2020
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your submission of Amendment/Modification materials for this project. The University of Maryland College Park (UMCP) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Prior to submission to the IRB Office, this project received scientific review from the departmental IRB Liaison.

This submission has received Expedited Review based on the applicable federal regulations.

This project has been determined to be a MINIMAL RISK project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of November 7, 2020.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Unless a consent waiver or alteration has been approved, Federal regulations require that each participant receives a copy of the consent document.

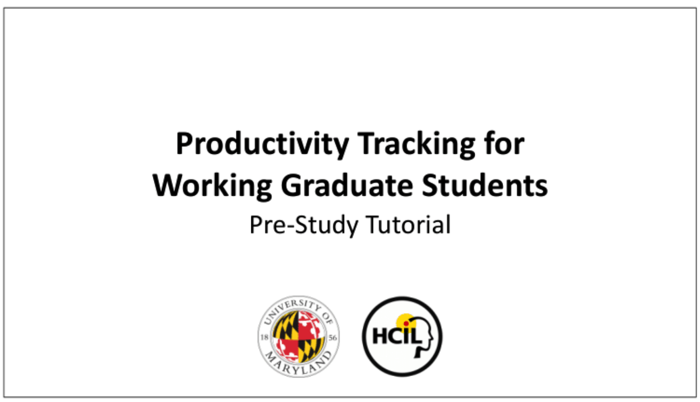
Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

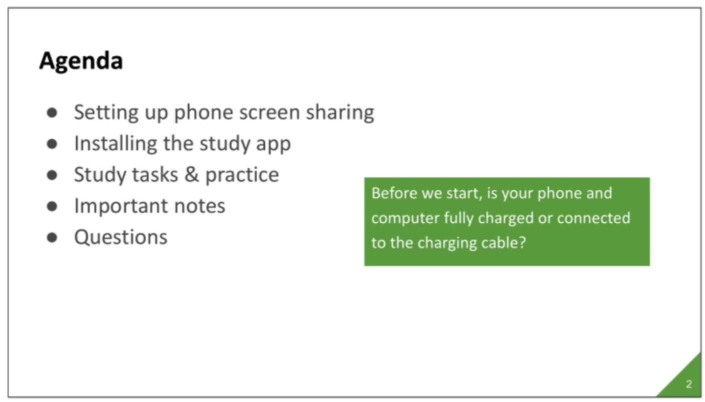
All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

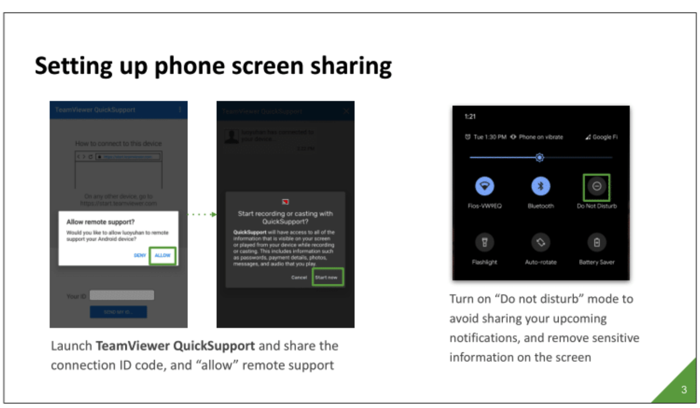
All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

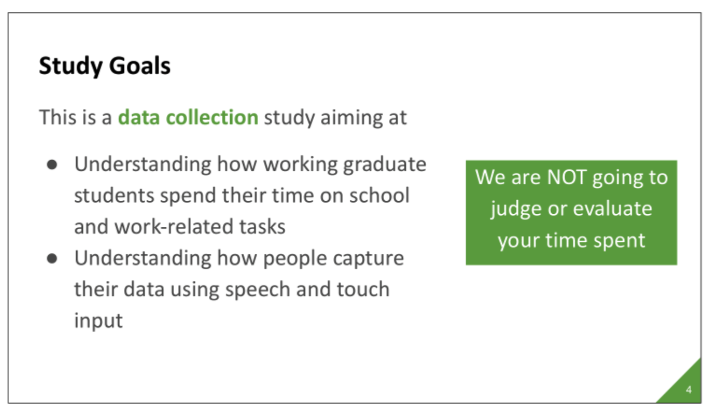
Appendix C: NoteWordy Study Material

C.1 Study Tutorial









Study Procedures

Pre-study tutorial (30-45 mins) → Data collection (14 days) → Debriefing interviews (Optional, 30 mins)

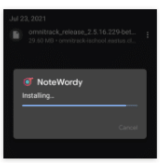
Receive $30 Amazon gift card

Receive additional $10 Amazon gift card

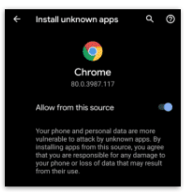
Installing the study app

Installing the study app

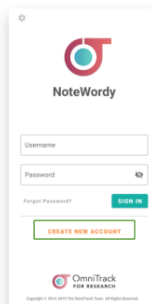
- Download the .APK file using the link in your email
- Tap the downloaded file to install the study app

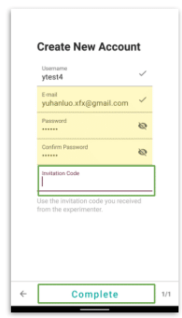


Make sure that your browser has the permission to install the app



Creating a study account



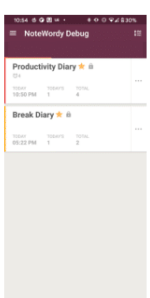


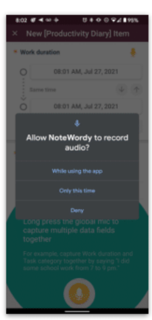
- Open the app and tap "Create New Account"
- Put your username, email, password
- Invite code: **UMD211700**
- Tap "Complete" at the bottom when you're done

Grant permissions

To make sure that the study app runs properly, you need to grant the following permissions:

- Allow audio recording permission for speech input
- Turn off the battery optimization (allow the app to run in the background to send you reminders)





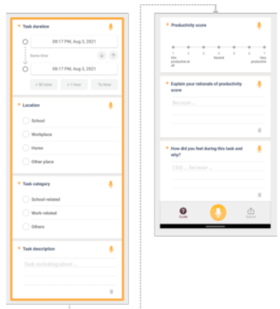
Study Tasks: "Productivity Diary" & "Break Diary"

"Productivity Diary": one task per entry

Task: a piece of work that you continuously work on during a period of time

- **Task duration**
  Start and end time of doing the task
- **Location**
  Where you are while doing the task
- **Task category**
  The task can be school-related, work-related, or others (e.g., daily chores, child care, volunteer)
- **Task description**
  Concrete description about the task (e.g., reading a paper, writing a report, debugging, data analysis)

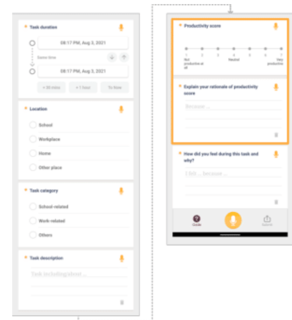
Q: What are your typical school-related and work-related tasks?



"Productivity Diary": one task per entry (Cont.)

Perceived productivity score and rationale

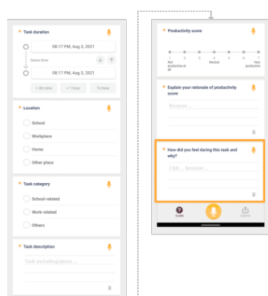
- **Productivity score**
  - "1" or "Not productive at all"
  - "2" or "Not productive"
  - "3" or "Somewhat not productive"
  - "4" or "Neutral"
  - "5" or "Somewhat productive"
  - "6" or "Productive"
  - "7" or "Very productive"
- **Rationale**
  Explain what makes you productive or not (e.g., figure out an important problem, complete a lot of work, feel satisfied with the outcome)



"Productivity Diary": one task per entry (Cont.)

How did you feel during the task and why

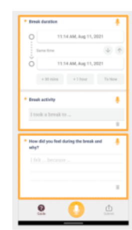
- Describe how you felt while doing this task (e.g., happy, excited, tired, upset, irritable), not your overall mood during that day
- Explain what made you feel in this way



"Break Diary": one break per entry

Break: **intentional break** that you take to refresh your mind

- **Break duration**
  Start and end time of the break
- **Break activity**
  What you did during the break (e.g., having coffee, taking a walk)



How did you feel during the break and why

- Describe how you felt while taking the break (not your overall mood during that day)
- Explain what made you feel in that way

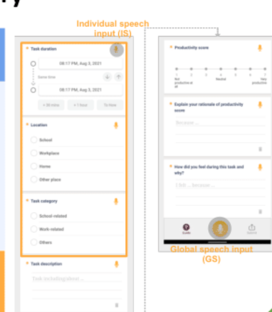
Q: What are some intentional break you take during work hours?

Study Tasks: Data capture with speech & touch input

Data capture in "Productivity Diary"

Data field	Individual speech input (IS)	Global speech input (GS)
Task duration (* remember to specify pm if it's afternoon)	- "*From 9 **pm** to 10 **pm**.*" - "*Started at 9 **pm** and **lasted** 1 hour.*"	Answer-by-answer: "*9 **pm** to 10 **pm**, **other places** and **work-related**.*" Natural language: "*I did some **school-related tasks** at **home from** 9 **to** 10 **pm**.*"
location	- "***Home**.*" - "*In the **school**.*"	
Task category	- "***job-related**.*" - "*It's a **work-related** task.*"	

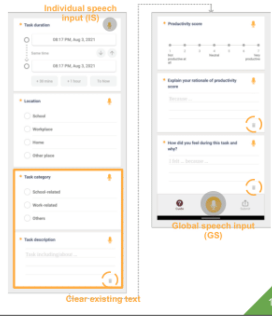
Practice time:
- Think about a task you complete today
- Use IS to capture a the *task duration*
- Use GS to *location* and *task category* together



Data capture in “Productivity Diary” (Cont.)

Data field	Individual speech input (IS)	Global speech input (GS)
Task category	- “***job-related***.” - “*It’s a* ***work-related*** *task.*”	“*I was doing some* ***school-related*** *tasks,* ***including*** *writing the final report of my class project, and reading some literature …*”
Task description (* say “clear” or use the delete button to clear existing text)	Anything that describes the task e.g., “*I was writing the final report of my class project.*”	

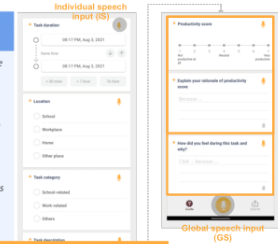
Practice time:
- Close the entry and reopen it, capture *task category* and *task description* together with GS
- Clear the text in *task description*, and capture the information again with IS



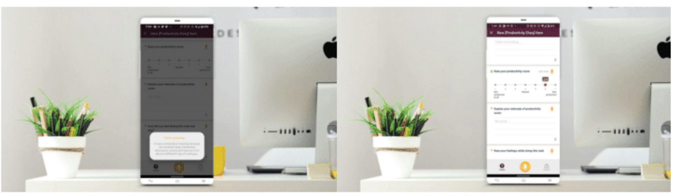
Data capture in “Productivity Diary” (Cont.)

Data field	Individual speech input (IS)	Global speech input (GS)
Productivity score	- “***1***” or “***Not productive at all***” - “***2***” or “***Not productive***” - “***3***” or “***Somewhat not productive***” - “***4***” or “***Neutral***” - “***5***” or “***Somewhat productive***” - “***6***” or “***Productive***” - “***7***” or “***Very productive***”	“*My* ***productivity*** *score is* ***5 because*** *I got the most important work done.* ***I felt*** *great* ***because*** *I got many interesting ideas while writing the report.*” “*I was* ***somewhat not productive****, because I still have a lot of things to finish.* ***I felt*** *okay* ***because*** *I know this is not an easy task.*”
Explain your rationale of productivity score	Anything that made you productive or not e.g., “*I got most of the work done.*”	
How did you feel during this task and why	Anything that describes and explains your feelings while doing the task e.g., “*I felt great because I enjoy writing, which made me think of many interesting ideas.*”	

Practice time:
- Use GS to capture *productivity score* and *rationale*



Demonstration: how to log your data with speech and touch input in the “Productivity Diary”



Log an entry using IS and GS

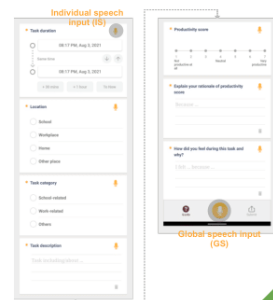
Clear existing text in the text entry

About the input modality

We encourage you to try out different ways of using speech input to capture your data, but it is not required to use speech all the time

- Use IS and GS as you see appropriate
- Feel free to choose between or combine speech and the original touch input in a way you feel comfortable

The speech recognition of NoteWordy is not perfect: you might encounter recognition errors, which can be fixed with another utterance or simply using touch input



Free Practice

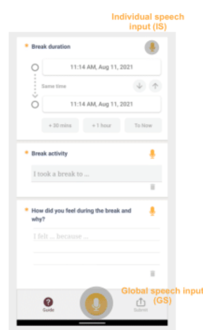
Think about another task you did today (school-related, work-related, or others), and log the information in the “Productivity Diary”.

Notes:
- Capture the information incrementally as you see appropriate. No need to follow the instructions
- The app may fail to recognize what you said--you can fix it with another utterance or touch input

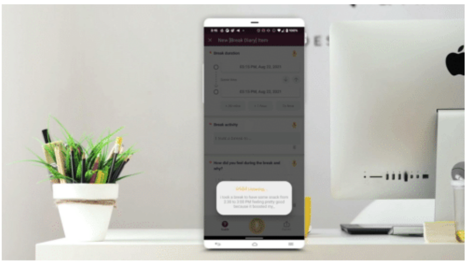
“Break Diary”

Data field	Individual speech input (IS)	Global speech input (GS)
Break duration	- “***From*** *10* ***to*** *10:30 am.*” - “***Started*** *at 3* ***pm*** *and* ***lasted*** *5 minutes.*”	“*I* ***had*** *some coffee* ***from*** *3* ***to*** *3:30* ***pm****, felt great because I was refreshed.*” “*My break was* ***from*** *3* ***to*** *3:30* ***pm****. I was doing some exercise, feeling great because it boosted my energy.*”
Break activity	Anything that describes what you did during the break e.g., “*I got some coffee.*”	
Why did you feel this way?	Anything that explains your feelings e.g., “*It’s was relaxing to take a walk.*”	

Practice time:
- Think about an intentional break you just take today (or yesterday).
- Capture all the data fields using GS



Demonstration: how to log your data with speech and touch input in the “Break Diary”



Free Practice

Think about the another break you took today (or yesterday), what activity you did, and how you felt. Then log the information in NoteWordy.

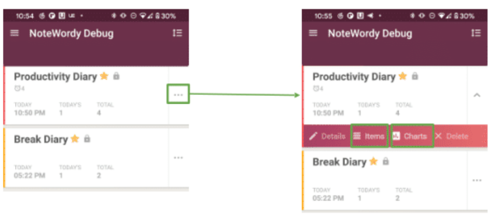
Notes:

- Capture the information incrementally as you see appropriate. No need to follow the instructions
- The app may fail to recognize what you said--you can fix it with another utterance or touch input

Review your logs

Tap the “three dots” icon on the diary you want to check

Tap the “items” to review your records



Important Notes

Study expectation: minimal requirement

Four time windows across a day:

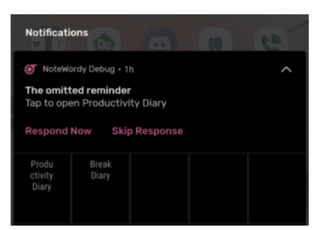
9 - 12 pm | **12 - 3 pm** | **3 - 6 pm** | **6 to 9 pm**

- “Productivity Diary”: **at least one entry of a representative task** in **three of the four** time windows per day
  - Three entries within one time window wouldn’t count
  - A task covers two time windows is fine, just make sure to log another two tasks covering the remaining time windows
- “Break Diary”: **At least one entry** of your **intentional break** per day
- You can also capture “other” tasks apart from school and work tasks (e.g., personal learning, chores)
- During the 14 days, you need to capture at least 12-day data (i.e., you can skip **2 days** at the most)

- You are encouraged to log as many entries as possible to cover all your tasks and breaks
- You will be dropped from the study if you cannot meet the minimal data capture requirement

Reminders

- 1st reminder: time to log morning tasks (12 pm)
- 2nd reminder: time to log afternoon tasks (3 pm)
- 3rd reminder: time to log late afternoon task (6 pm)
- 4th reminder: time to log evening tasks (9 pm)



Compensation

- $ 30 at the end of the two-week data collection
- Additional $10 if you opt in for the debriefing interview

- Compensation will be provided in the form of an Amazon Gift Card.
- If you are dropped from the study, you will NOT receive the study compensation.

Next steps

- The data collection starts **tomorrow** (from 9/15 - 9/28/2021)
- You are encouraged to log all your representative tasks and intentional breaks at your best
- We will contact you to ask your interest in opting in the debriefing interview at the end of the study on 9/28/2021
- We will use TeamViewer QuickSupport again during the debriefing interview
- You will receive the study compensation once we confirm you have completed all the study tasks

Questions?

C.2 Interview Questions

Q1. To start, how was your experience of tracking your tasks and breaks in NoteWordy over the past two weeks?

Q2. How was your experience with the input modalities including manual touch, speech, and the combination of the two?

- When you log your data, how do you decide whether to use touch or speech input? Why is that?
- When you use speech input, how often did recognition errors occur? What would you do in those situations?

Q3. When you use speech input, how do you decide whether to use the global speech icon at the bottom center or the individual speech icon at each data field? Why is that?

- I noticed that sometimes you would combine both speech and touch input in one entry (refer to a specific entry), can you recall your experience of using both modalities together?
- When do you think the global speech input icon is most useful? Why?
- When do you think the individual speech input icon is most useful? Why?
- When do you think touch input is most useful? Why?

Q4.What was the experience like when you capture data in the Productivity Diary vs in the Break Dairy?

- Any differences regarding the input choices in these two diaries?
- Have you used the delete icon in text entries? Can you tell me more about it?

Q5. Where were you most of the time when you were capturing the data?

- Did others' present influence your choice between speech vs. text input? Why?
- Do you have privacy concerns regarding speech input?

Q6. Regardless of this app, what was your overall experience with speech input?

- Is there anything specific about capturing data using NoteWordy? (use it in front of others, error tolerance)
- Did you learn anything new from using speech and touch input in this study?

Q7. What did you learn from the two-week data collection on your productivity and break data?

- What kinds of feedback do you hope to receive from the data you logged in your Productivity and Break Diary, regardless of technical constraints?

Q8. Would you like to keep using NoteWordy in the future, or recommend it to your friends who are also working graduate students? Why or why not?

Q9. Anything you want to share with us?

C.3 IRB Approval letter

UNIVERSITY OF MARYLAND
INSTITUTIONAL REVIEW BOARD

1204 Marie Mount Hall
College Park, MD 20742-5125
TEL 301.405.4212
FAX 301.314.1475
irb@umd.edu
www.umresearch.umd.edu/IRB

DATE:	August 5, 2021
TO:	Eun Kyoung Choe
FROM:	University of Maryland College Park (UMCP) IRB
PROJECT TITLE:	[1132164-23] Supporting personal data tracking and data exploration
REFERENCE #:	
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	August 5, 2021
EXPIRATION DATE:	November 7, 2021
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *7*

Thank you for your submission of Amendment/Modification materials for this project. The University of Maryland College Park (UMCP) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Prior to submission to the IRB Office, this project received scientific review from the departmental IRB Liaison.

This submission has received Expedited Review based on the applicable federal regulations.

This project has been determined to be a MINIMAL RISK project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of November 7, 2021.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Unless a consent waiver or alteration has been approved, Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

Appendix D: TandemTrack Study Material

D.1 Study Tutorial

D.1.1 Tutorial for M Group

TandemTrack Study Tutorial

Yuhan Luo 03/14/2019
College of Information Studies
University of Maryland

1

Agenda

- The goal of the study
- Study Procedure
- Exercise with the Study App, TandemTrack
- Important notes
- Questions

2

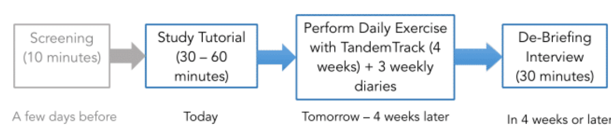
The goal of the study

To help you create a consistent daily exercise habit (i.e., sit-up & push-up) with the study App, TandemTrack.

Icons made by Darius Dan from Flaticon is licensed by CC 3.0 BY.

TandemTrack Study Tutorial 3

Study Procedure



TandemTrack Study Tutorial 4

Daily Exercise Task

- A session of sit-ups/push-ups **alternatively**

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	...
Sit-up	Push-up	Sit-up	Push-up	Sit-up	Push-up	Sit-up	...

- One session lasts 3 minutes and 30 seconds：

- We encourage you to complete the 3 sets in one time. Only completing 1 or 2 sets does not count as a completion of your daily task.

Weekly Diary

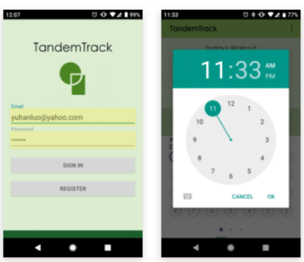
A Google Form at the end of each week:

- How was your experience with TandemTrack over the past week?
- Did you experience any technical difficulties?
- Is there anything you want to share with us?

Briefly respond to each question in one or two sentence and get back to us in the next 12 hours.

You will receive 3 weekly diaries during the study, in the last (3rd) weekly diary, we will schedule the last de-briefing interview with you

Exercise with TandemTrack App:
Log in & set reminder

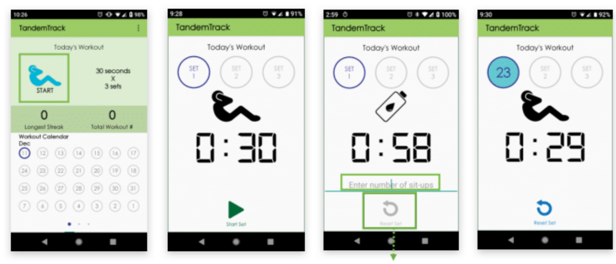


Log in

Set a time to receive daily exercise reminder

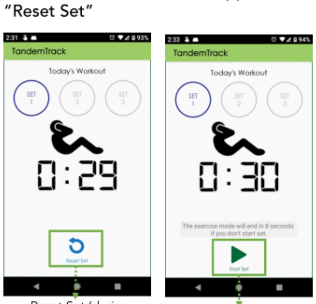
- We encourage you to start exercising upon receiving the reminder, but you're not required to do so
- The reminder time cannot be changed
- Think about a time when you would like to receive the daily exercise reminder
- You will receive the reminder starting tomorrow
- If you have already completed the exercise before the reminder time, you will not receive the reminder from the App

Exercise with TandemTrack App:
Exercise flow



Cannot press the button when it is greyed out

Exercise with TandemTrack App:
"Reset Set"

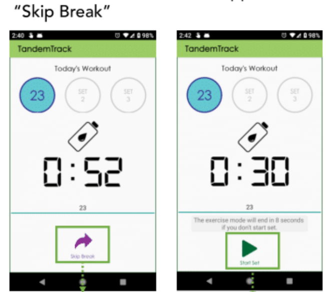


Reset Set (during exercise)

Start Set

- You can reset the current set during exercise
- When you are ready, press "Start Set"
- **If you don't press "Start Set" within 90 seconds, the exercise mode will exit and your data will not be stored**

Exercise with TandemTrack App:
"Skip Break"

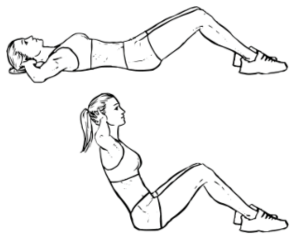


Skip Break (during break)

Start Set

- **During the break,** you can press the "Skip Break" button **after** you enter the exercise number.
- After skipping a break, you need to press "Start Set" to proceed to the next set
- **If you don't press "Start Set" within 90 seconds, the exercise mode will exit and your data will be not stored**

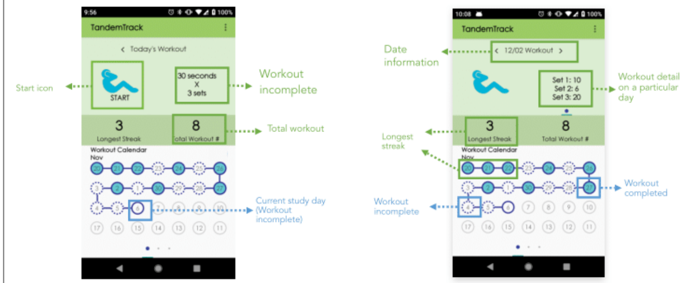
Exercise with TandemTrack App: Sit-ups



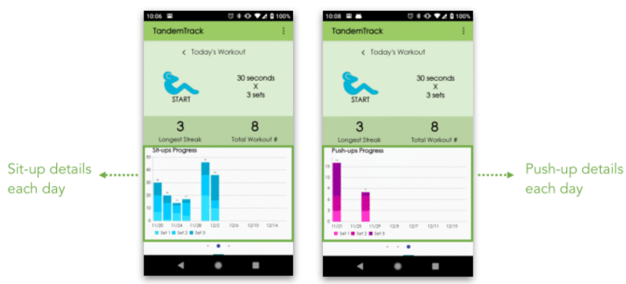
- Remember to count the number of sit-ups you complete in each set
- Be aware of the transition of background sounds

Images from workoutlab.com

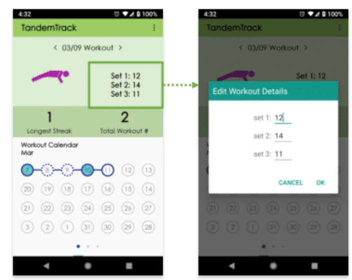
Exercise with TandemTrack App: How to interpret the visualization?



Exercise with TandemTrack App: How to interpret the visualization?

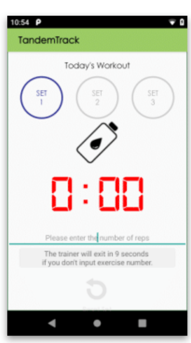


Exercise with TandemTrack App: Edit exercise details



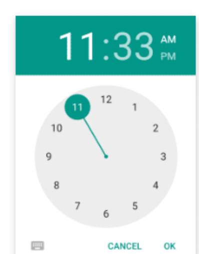
Important notes

- You are encouraged to complete the 3 sets within 3 minutes and 30 seconds by following the exercise guidance. Please don't reset set unless you really need to
- If you do not start set after resetting/ skipping break or do not input exercise number on time, TandemTrack will time out.



Important notes

- Your first day of study is **tomorrow,** so **don't** do exercise today
- You **cannot** change the time of daily reminder once you set it
- We encourage you to start exercising upon receiving the reminder, but this is **not** required



D.1.2 Tutorial for MA Group

TandemTrack
Study Tutorial

Yuhan Luo 02/22/2019
College of Information Studies
University of Maryland

1

Agenda

- The goal of the study
- Study Procedure
- Exercise with the Study App, TandemTrack
- Exercise with the Alexa skill version of TandemTrack
- Important notes
- Questions

2

The goal of the study

To help you create a **consistent daily exercise habit** (i.e., sit-up & push-up) with the study App, TandemTrack.

Icons made by Darius Dan from Flaticon is licensed by CC 3.0 BY.

TandemTrack Study Tutorial 3

Study Procedure

Screening (10 minutes) → Study Tutorial (45 – 60 minutes) → Perform Daily Exercise with TandemTrack (4 weeks) + 3 weekly diaries → De-Briefing Interview (30 minutes)

A few days before | Today | Tomorrow – 4 weeks later | In 4 weeks or later

TandemTrack Study Tutorial 4

Daily Exercise Task

- A session of sit-ups/push-ups **alternatively**

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	...
Sit-up	Push-up	Sit-up	Push-up	Sit-up	Push-up	Sit-up	...

- One session lasts 3 minutes and 30 seconds:

- We encourage you to complete the 3 sets in one time. Only completing 1 or 2 sets does not count as a completion of your daily task.

Weekly Diary

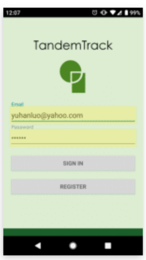
A Google Form at the end of each week:

- How was your experience with TandemTrack over the past week?
- Did you experience any technical difficulties?
- Is there anything you want to share with us?

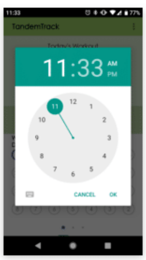
Briefly respond to each question in one or two sentence and get back to us in the next 12 hours.

You will receive 3 weekly diaries during the study, in the last (3rd) weekly diary, we will schedule the last de-briefing interview with you

Exercise with TandemTrack App:
Log in & set reminder



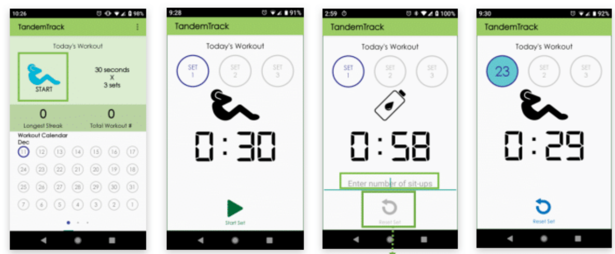
Log in



Set a time to receive daily exercise reminder

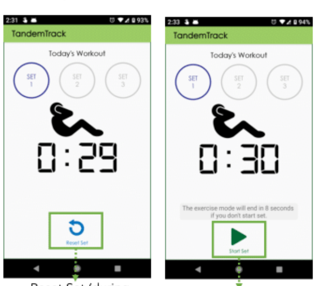
- We encourage you to start exercising upon receiving the reminder, but you're not required to do so
- The reminder time cannot be changed
- Think about a time when you would like receive the daily exercise reminder
- You will receive the reminder starting tomorrow
- If you have already completed the exercise before the reminder time, you will not receive the reminder from the App

Exercise with TandemTrack App:
Exercise flow



Cannot press the button when it is greyed out

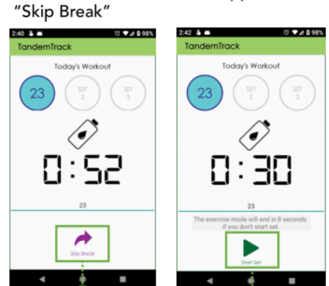
Exercise with TandemTrack App:
"Reset Set"



Reset Set (during exercise) Start Set

- You can reset the current set during exercise
- When you are ready, press "Start Set"
- **If you don't press "Start Set" within 90 seconds, the exercise mode will exit and your data will not be stored**

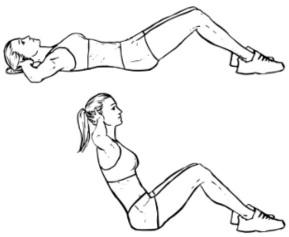
Exercise with TandemTrack App:
"Skip Break"



Skip Break (during break) Start Set

- **During the break,** you can press the "Skip Break" button **after** you enter the exercise number.
- After skipping a break, you need to press "Start Set" to proceed to the next set
- **If you don't press "Start Set" within 90 seconds, the exercise mode will exit and your data will not be stored**

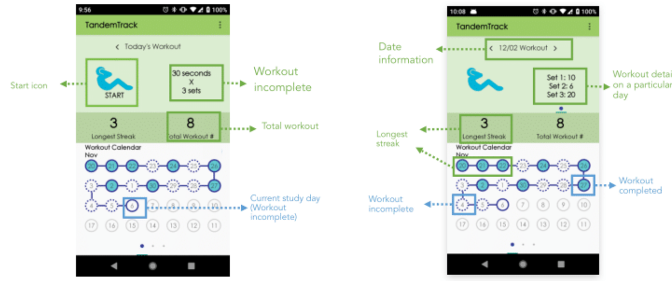
Exercise with TandemTrack App: Sit-ups



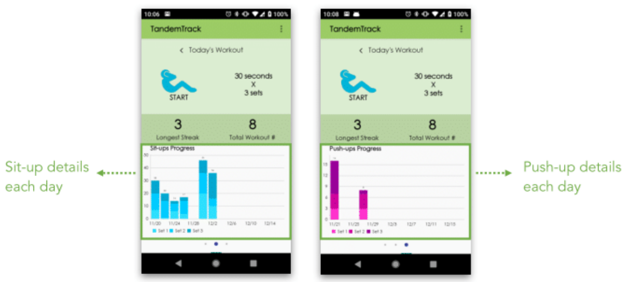
- Remember to count the number of sit-ups you complete in each set
- Be aware of the transition of background sounds

Images from workoutlab.com

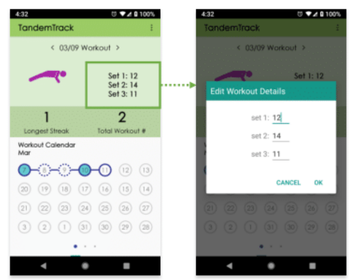
Exercise with TandemTrack App: How to interpret the visualization?



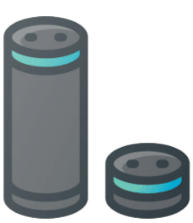
Exercise with TandemTrack App: How to interpret the visualization?



Exercise with TandemTrack App: Edit exercise details

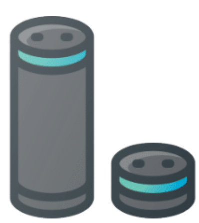


Exercise with TandemTrack Skill: What can it do?



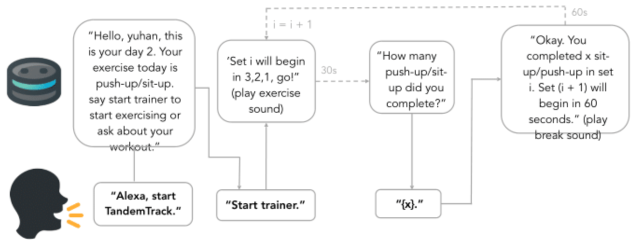
- The TandemTrack skill shares the same features with the TandemTrack App
  - Exercise guidance
  - Exercise data capture
  - Feedback of exercise data
  - Reminder
- The database is shared with TandemTrack App

Exercise with TandemTrack Skill: Set up the device and invoke the skill

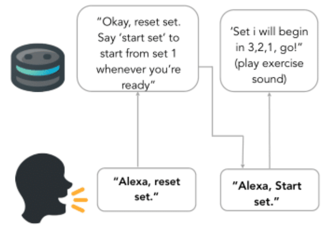


- You will need to set up the device at home with the study account, following the instruction (User Manual, pg9).
- Set a daily exercise reminder
- Disable notification from you Alexa App
- The reminder is not tied to the TandemTrack skill

Exercise with TandemTrack Skill: Start trainer

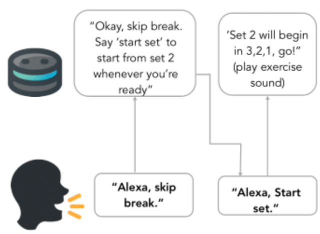


Exercise with TandemTrack Skill: "Reset set"



- You can only say "reset set" during exercise, but not during break
- You need to invoke the skill by saying "Alexa" before resetting a set
- If you don't say "Alexa, start set" after 90 seconds since your said "Alexa, reset set", Alexa will time out.

Exercise with TandemTrack Skill: "Skip break"



- You can only say "skip break" during break, but not during exercise
- You need to invoke the skill by saying "Alexa" before skipping a break
- If you don't say "Alexa, start set" after 90 seconds since your said "Alexa, skip break", Alexa will time out.

Exercise with TandemTrack Skill: Start trainer

Push-ups

- Remember the number of push-ups you complete in each set
- Be aware of the transition of background sounds

Images from workoutlab.com

Exercise with TandemTrack Skill: Receive feedback

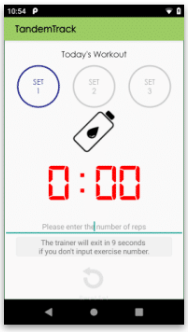
Intent	Example utterances
Ask for workout summary	"workout summary."
Ask for streak	"What is my current/longest streak?"
Ask for total sessions of workout	"How many push-ups/sit-ups/workout did I do in total?"
Ask for average number of workout per session	"Average push-ups/sit-ups/workout" "How many push-ups/sit-ups/workout did I do in average?"
Ask for minimal workout per session	"Minimal push-ups/sit-ups/workout" "How many push-ups/sit-ups/workout did I do in minimum?"
Ask for maximal workout per session	"What is my maximal push-ups/sit-ups/workout?"
Ask for workout on a particular day	"How many workout/push-ups/sit-ups did I do today/ yesterday/November 1st?"

Important notes

- You can interact with the TandemTrack App or skill as the way you like
- Your data will be synced on both devices
- Please **don't** start trainer on TandemTrack Skill and App at the same time

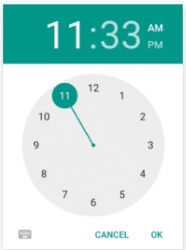
Important notes

- You are encouraged to complete the 3 sets within 3 minutes and 30 seconds by following the exercise guidance. Please don't reset set unless you really need to
- If you do not start set after resetting/ skipping break or do not input exercise number on time, TandemTrack will time out.



Important notes

- Your first day of study is **tomorrow,** so **don't** do exercise today
- You **cannot** change the time of daily reminder once you set it
- We encourage you to start exercising upon receiving the reminder, but this is **not** required



D.2 Weekly Diary

Q1. What was your experience like with TandemTrack over the past week?

Q2. Did you experience any technical problems with TandemTrack?

Q3. Is there any thing you want to share with us?

D.3 Interview Questions

D.3.1 Questions For Both Groups

Q1. How was your experience with TandemTrack over the past four weeks?

Q2. Where did you usually do exercise?

- What is your living environment? Do you have roommates?

Q3. When did you usually do exercise?

- Is your exercise time close to the reminder time or not?

Q4. (if there are missed days) I noticed that you missed your exercise on day x and y, can you recall the reasons why you missed the exercise?

- What do you think is the most challenging part in completing the exercise sessions?

Q5. (If there are no missed days) I noticed that you completed all the exercise every day, which is amazing. How did you make it?

- What is working in keeping you doing exercise?

- Were there any challenges for you to complete the exercise goal?

Q6. How did you feel about the exercise guidance?

- How easy or difficult to follow the guidance?
- What did you do during the break?

Q7. How easy or difficult was it to capture your exercise data?

- Was there any cases you entered wrong numbers?
- is there any cases you modify your exercise number?

Q8. How often did you check your exercise feedback?

- Why or why not do you check your exercise feedback?

Q9. Have you ever used any other fitness app, or self-tracking apps?

- Can you give some examples?

Q10. Anything you hope TandemTrack could support but it didn't?

Q11. Are you still doing sit-ups/push-ups after the study is over? Are you considering doing this in the future?

D.3.2 Additional Questions For MA Group

Q12. Where was the Echo device located at your home?

- Is that consistent?
- How far were you usually from the device?
- Can you hear the exercise reminder from the device most of the time?

Q13.How did you typically use the two versions of TandemTrack?

- When you do exercise with TandemTrack, how did you decide whether to use mobile phone or Echo dot?
- Is there any case you use Alexa and mobile app of TandemTrack together?
- Did you encounter any challenges or difficulties in using any of the version?
- In general, which version did you use more?

Q14. Which version of TandemTrack do you prefer using and why?

- In what cases did you like to use the Alexa version more than the mobile version, and why?
- In what cases did you like to use the mobile version more than the Alexa version, and why?
- In general, what motivate you to use the Echo device?

Q15. What are the challenges/barriers in interaction with the Echo device?

- Any issues with speech recognition?

Q16. Did you use the Alexa for other purposes besides TandemTrack?

- Can you give some examples?

Q17. Anything you would like to share with us?

D.4 IRB Approval letter

UNIVERSITY OF MARYLAND
INSTITUTIONAL REVIEW BOARD

1204 Marie Mount Hall
College Park, MD 20742-5125
TEL 301.405.4212
FAX 301.314.1475
irb@umd.edu
www.umresearch.umd.edu/IRB

DATE:	March 29, 2019
TO:	Eun Kyoung Choe
FROM:	University of Maryland College Park (UMCP) IRB
PROJECT TITLE:	[1132164-9] Supporting personal data tracking and data exploration
REFERENCE #:	
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	March 29, 2019
EXPIRATION DATE:	November 7, 2019
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your submission of Amendment/Modification materials for this project. The University of Maryland College Park (UMCP) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Prior to submission to the IRB Office, this project received scientific review from the departmental IRB Liaison.

This submission has received Expedited Review based on the applicable federal regulations.

This project has been determined to be a MINIMAL RISK project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of November 7, 2019.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Unless a consent waiver or alteration has been approved, Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

Bibliography

[1] Rosemery O Nelson. Assessment and therapeutic functions of self-monitoring. In *Progress in behavior modification*, volume 5, pages 263–308. Elsevier, 1977.

[2] William J Korotitsch and Rosemery O Nelson-Gray. An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, 11(4):415, 1999.

[3] 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 2014 CHI Conference on Human Factors in Computing Systems*, pages 1143–1152, 2014.

[4] 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*, pages 557–566, 2010.

[5] Digital Health Consumer Adoption Report 2019. https://rockhealth.com/insights/digital-health-consumer-adoption-report-2019/. Accessed: 2022-04-11.

[6] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Aaai*, volume 5, pages 1541–1546. Pittsburgh, PA, 2005.

[7] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L Cox. Flexible and mindful self-tracking: Design implications from paper bullet journals. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[8] Johnna Blair, Yuhan Luo, Ning F Ma, Sooyeon Lee, and Eun Kyoung Choe. Onenote meal: A photo-based diary study for reflective meal tracking. In *AMIA Annual Symposium Proceedings*, volume 2018, page 252. American Medical Informatics Association, 2018.

[9] Haining Zhu, Yuhan Luo, and Eun Kyoung Choe. Making space for the quality care: Opportunities for technology in cognitive behavioral therapy for insomnia. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5773–5786, 2017.

[10] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A Munson. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 3(1):1–27, 2019.

[11] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. Sleeptight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 121–132, 2015.

[12] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, pages 1159–1162, 2015.

[13] Xiaochen Huang and Emmanuel Agu. A speech-based mobile app for restaurant order recognition and low-burden diet tracking. In *ICSH*, pages 333–339. Springer, 2015.

[14] Barbara L Chalfonte, Robert S Fish, and Robert E Kraut. Expressive richness: a comparison of speech and text as media for revision. In *Proceedings of the 1991 CHI Conference on Human Factors in Computing Systems*, pages 21–26, 1991.

[15] Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James A Landay. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–23, 2018.

[16] 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*, pages 1–17, 2021.

[17] Melanie Revilla, Mick P Couper, Oriol J Bosch, and Marc Asensio. Testing the use of voice input in a smartphone web survey. *Social Science Computer Review*, 38(2):207–224, 2020.

[18] Amazon Echo. https://www.amazon.com/echo, 2022. Accessed: 2022-04-11.

[19] Eugene Charniak and Mark Johnson. Edit detection and parsing for transcribed speech. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*, 2001.

[20] Christine Murad, Cosmin Munteanu, Leigh Clark, and Benjamin R Cowan. Design guidelines for hands-free speech interaction. In *MobileHCI 2018*, pages 269–276.

[21] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. Omnitrack: A flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–28, 2017.

[22] Ivan Robertson and Cary Cooper. *Well-being: Productivity and happiness at work*. Springer, 2011.

[23] Gloria Mark, Mary Czerwinski, Shamsi Iqbal, and Paul Johns. Workplace indicators of mood: Behavioral and cognitive correlates of mood among information workers. In *Proceedings of the 6th International Conference on Digital Health Conference*, pages 29–36, 2016.

[24] Judith S Hurwitz, Alan Nugent, Fern Halper, and Marcia Kaufman. *Big data for dummies*. John Wiley & Sons, 2013.

[25] Developing a healthcare app in 2021: What do patients really want? https://www.imaginovation.net/blog/developing-a-mobile-health-app-what-patients-really-want/. Accessed: 2022-04-11.

[26] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 272–283, 2016.

[27] Elizabeth L Murnane, Dan Cosley, Pamara Chang, Shion Guha, Ellen Frank, Geri Gay, and Mark Matthews. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association*, 23(3):477–484, 2016.

[28] Steven C Hayes and Norman Cavior. Multiple tracking and the reactivity of self-monitoring: I. negative behaviors. *Behavior Therapy*, 8(5):819–831, 1977.

[29] Albert J Stunkard and Michael J Mahoney. Behavioral treatment of the eating disorders. *Handbook of behavior modification and behavior therapy/Harold Leitenberg*, 1976.

[30] Beate Pfeiffer, Trent Stellingwerff, Adrian B Hodgson, Rebecca Randell, Klaus Pöttgen, Peter Res, and Asker E Jeukendrup. Nutritional intake and gastrointestinal problems during competitive endurance events. *Medicine & Science in Sports & Exercise*, 44(2):344–351, 2012.

[31] Yunsheng Ma, Elizabeth R Bertone, Edward J Stanek III, George W Reed, James R Hebert, Nancy L Cohen, Philip A Merriam, and Ira S Ockene. Association between

eating patterns and obesity in a free-living us adult population. *American journal of epidemiology*, 158(1):85–92, 2003.

[32] Matthew D Mitchell, Philip Gehrman, Michael Perlis, and Craig A Umscheid. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. *BMC family practice*, 13(1):1–11, 2012.

[33] Walter C Willett, Laura Sampson, Meir J Stampfer, Bernard Rosner, Christopher Bain, Jelia Witschi, Charles H Hennekens, and Frank E Speizer. Reproducibility and validity of a semiquantitative food frequency questionnaire. *American journal of epidemiology*, 122(1):51–65, 1985.

[34] Patricia B Crawford, Eva Obarzanek, John Morrison, and ZI Sabry. Comparative advantage of 3-day food records over 24-hour recall and 5-day food frequency validated by observation of 9-and 10-year-old girls. *Journal of the American Dietetic Association*, 94(6):626–630, 1994.

[35] Gladys Block. A review of validations of dietary assessment methods. *American journal of epidemiology*, 115(4):492–505, 1982.

[36] Rosa M Ortega, Carmen Pérez-Rodrigo, and Ana M López-Sobaler. Dietary assessment methods: dietary records. *Nutricion hospitalaria*, 31(3), 2015.

[37] Elina Mattila, Juha Pärkkä, Marion Hermersdorf, Jussi Kaasinen, Janne Vainio, Kai Samposalo, Juho Merilahti, Juha Kolari, Minna Kulju, Raimo Lappalainen, et al. Mobile diary for wellness management—results on usage and usability in two user studies. *IEEE Transactions on information technology in biomedicine*, 12(4):501–512, 2008.

[38] Kanitthika Kaewkannate and Soochan Kim. A comparison of wearable fitness devices. *BMC public health*, 16(1):433, 2016.

[39] Susannah Fox and Maeve Duggan. *Tracking for health*. Pew Research Center's Internet & American Life Project, 2013.

[40] Haining Zhu, Joanna Colgan, Madhu Reddy, and Eun Kyoung Choe. Sharing patient-generated data in clinical practices: an interview study. In *AMIA Annual Symposium Proceedings*, volume 2016, page 1303. American Medical Informatics Association, 2016.

[41] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, Sean A Munson, et al. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of medical Internet research*, 17(8):e4364, 2015.

[42] Jasmine K Zia, C-F Chung, Jessica Schroeder, Sean A Munson, Julie A Kientz, James Fogarty, Elizabeth Bales, Jeanette M Schenk, and Margaret M Heitkemper.

The feasibility, usability, and clinical utility of traditional paper food and symptom journals for patients with irritable bowel syndrome. *Neurogastroenterology & Motility*, 29(2):e12935, 2017.

[43] Alan E Kazdin. Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. *Journal of consulting and clinical psychology*, 42(5):704, 1974.

[44] Ruth Baumgart and Lukas Wiewiorra. The role of self-control in self-tracking. 2016.

[45] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems*, pages 1797–1806, 2008.

[46] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*, pages 261–278. Springer, 2006.

[47] Jeni Paay, Jesper Kjeldskov, Mikael B Skov, Nirojan Srikandarajah, and Umachanger Brinthaparan. Quittylink: Using smartphones for personal counseling to help people quit smoking. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 98–104, 2015.

[48] Kelly D Brownell, G Alan Marlatt, Edward Lichtenstein, and G Terence Wilson. Understanding and preventing relapse. *American psychologist*, 41(7):765, 1986.

[49] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. How to evaluate technologies for health behavior change in hci research. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3063–3072, 2011.

[50] Fitbit. https://www.fitbit.com/, 2022. Accessed: 2022-04-11.

[51] Myfooddiary. https://www.myfooddiary.com/. Accessed: 2022-04-11.

[52] Under Armour, Inc. MyFitnessPal, 2022. Accessed: 2022-04-11.

[53] Matthew L Lee and Anind K Dey. Reflecting on pills and phone use: supporting awareness of functional abilities for older adults. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems*, pages 2095–2104, 2011.

[54] Google Fit. https://www.fitbit.com/, 2022. Accessed: 2022-04-11.

[55] Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H Nguyen. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 635–646, 2015.

[56] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. Lullaby: a capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 226–234, 2012.

[57] Xianghua Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. Data engagement reconsidered: A study of automatic stress tracking technology in use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.

[58] Daniel A Epstein, Daniel Avrahami, and Jacob T Biehl. Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 673–684, 2016.

[59] Bite AI, Inc. Bitesnap. https://getbitesnap.com/, 2022. Accessed: 2022-04-11.

[60] Journify. https://journify.co/. Accessed: 2022-04-11.

[61] Genesant Technologies, Inc. Talk-to-Track, 2022. Accessed: 2022-04-11.

[62] Jakob Tholander and Maria Normark. Crafting personal information-resistance, imperfection, and self-creation in bullet journaling. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[63] Mark Matthews and Gavin Doherty. In the mood: engaging teenagers in psychotherapy using mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2947–2956, 2011.

[64] Parastoo Abtahi, Victoria Ding, Anna C Yang, Tommy Bruzzese, Alyssa B Romanos, Elizabeth L Murnane, Sean Follmer, and James A Landay. Understanding physical practices and the role of technology in manual self-tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–24, 2020.

[65] Judy Kopp. Self-monitoring: A literature review of research and practice. In *Social Work Research and Abstracts*, volume 24, pages 8–20. Oxford University Press, 1988.

[66] Jessica Schroeder, Chia-Fang Chung, Daniel A Epstein, Ravi Karkar, Adele Parsons, Natalia Murinova, James Fogarty, and Sean A Munson. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pages 135–148, 2018.

[67] Xinghui Yan, Katy Madier, Sun Young Park, and Mark Newman. Towards low-burden in-situ self-reporting: A design space exploration. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, pages 337–346, 2019.

[68] Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A Munson. Examining opportunities for goal-directed self-tracking to support chronic condition management. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 3(4):1–26, 2019.

[69] Nicole Freene, Gordon Waddington, Wendy Chesworth, Rachel Davey, and Tom Cochrane. Validating two self-report physical activity measures in middle-aged adults completing a group exercise or home-based physical activity program. *Journal of science and medicine in sport*, 17(6):611–616, 2014.

[70] Mette Aadahl and Torben Jørgensen. Validation of a new self-report instrument for measuring physical activity. *Medicine and science in sports and exercise*, 35(7):1196–1202, 2003.

[71] Xiaoyi Zhang, Laura R Pina, and James Fogarty. Examining unlock journaling with diaries and reminders for in situ self-report in health and wellness. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5658–5664, 2016.

[72] Eun Kyoung Choe. *Designing self-monitoring technology to promote data capture and reflection*. PhD thesis, 2014.

[73] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A Epstein, Felicia Cordeiro, Matthew Kay, Gregory D Abowd, Tanzeem Choudhury, James Fogarty, et al. Semi-automated tracking: a balanced approach for self-monitoring applications. *IEEE Pervasive Computing*, 16(1):74–84, 2017.

[74] Young-Ho Kim. *Designing Flexible Self-Tracking Technologies fo Enhancing In Situ Data Collection Capability*. PhD thesis, 2019.

[75] 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.

[76] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E McCullough, and Rashid Ansari. Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3):171–193, 2002.

[77] Dong Yu and Li Deng. *Automatic Speech Recognition.* Springer, 2016.

[78] Allyson Ettinger, Sudha Rao, Hal Daumé III, and Emily M Bender. Towards linguistically generalizable nlp systems: A workshop and shared task. *arXiv preprint arXiv:1711.01505*, 2017.

[79] Otter.ai. https://otter.ai/, 2022. Accessed: 2022-04-11.

[80] Anuj Kumar, Tim Paek, and Bongshin Lee. Voice typing: a new speech interaction model for dictation on touchscreen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2277–2286, 2012.

[81] 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.

[82] Amazon Alexa. https://alexa.amazon.com/, 2022. Accessed: 2022-04-11.

[83] Google, Inc. Google assistant, 2022. Accessed: 2022-04-11.

[84] Apple, Inc. Siri, 2022. Accessed: 2022-04-11.

[85] Google. 5 ways voice assistance is shaping consumer behavior. https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/voice-assistance-consumer-experience/, 2019. Accessed: 2022-04-11.

[86] Marika De Bruijne and Arnaud Wijnant. Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey. *Social Science Computer Review*, 31(4):482–504, 2013.

[87] Michael F Schober, Frederick G Conrad, Christopher Antoun, Patrick Ehlen, Stefanie Fail, Andrew L Hupp, Michael Johnston, Lucas Vickers, H Yanna Yan, and Chan Zhang. Precision and disclosure in text and voice interviews on smartphones. *PloS one*, 10(6):e0128337, 2015.

[88] Frederick G Conrad, Michael F Schober, Christopher Antoun, H Yanna Yan, Andrew L Hupp, Michael Johnston, Patrick Ehlen, Lucas Vickers, and Chan Zhang. Respondent mode choice in a smartphone survey. *Public Opinion Quarterly*, 81(S1):307–337, 2017.

[89] Somani Patnaik, Emma Brunskill, and William Thies. Evaluating the accuracy of data collection on mobile phones: A study of forms, sms, and voice. In *2009 International Conference on Information and Communication Technologies and Development (ICTD)*, pages 74–84. IEEE, 2009.

[90] Yixuan Zhang and Andrea G Parker. Eat4thought: A design of food journaling. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[91] Yuhan Luo, Bongshin Lee, and Eun Kyoung Choe. Tandemtrack: shaping consistent exercise experience by complementing a mobile app with a smart speaker. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2020.

[92] ID Hill. Natural language versus computer language. *Designing for human-computer communication*, pages 55–72, 1983.

[93] Cosmin Munteanu, Matt Jones, Sharon Oviatt, Stephen Brewster, Gerald Penn, Steve Whittaker, Nitendra Rajput, and Amit Nanavati. We need to talk: Hci and the delicate topic of spoken language interaction. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2459–2464. 2013.

[94] Asbjørn Følstad and Petter Bae Brandtzæg. Chatbots and the new world of hci. *interactions*, 24(4):38–42, 2017.

[95] Fatma Őzcan, Abdul Quamar, Jaydeep Sen, Chuan Lei, and Vasilis Efthymiou. State of the art and open challenges in natural language interfaces to data. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pages 2629–2636, 2020.

[96] Dayne Freitag. Machine learning for information extraction in informal domains. *Machine learning*, 39(2):169–202, 2000.

[97] Jakub Piskorski and Roman Yangarber. Information extraction: Past, present and future. In *Multi-source, multilingual information extraction and summarization*, pages 23–49. Springer, 2013.

[98] Day One, 2022. Accessed: 2022-04-11.

[99] Midnight Plan, Inc. MurMur, 2022. Accessed: 2022-04-11.

[100] Mandy Korpusik, Salima Taylor, Sai Krupa Das, Cheryl Gilhooly, Susan Roberts, and James Glass. A food logging system for ios with natural spoken language meal descriptions (p21-009-19). *Current developments in nutrition*, 3(Supplement_1):nzz041–P21, 2019.

[101] Lucas M Silva and Daniel A Epstein. Investigating preferred food description practices in digital food journaling. In *Proceedings of the 2021 Conference on Designing Interactive System*. ACM, 2021.

[102] Helaine RH Rockett, Anne M Wolf, and Graham A Colditz. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *Journal of the American Dietetic Association*, 95(3):336–340, 1995.

[103] Ruth E Patterson, Alan R Kristal, Lesley Fels Tinker, Rachel A Carter, Mary Pat Bolton, and Tanya Agurs-Collins. Measurement characteristics of the women's health initiative food frequency questionnaire. *Annals of epidemiology*, 9(3):178–187, 1999.

[104] FitNow, Inc. Lose It! – Weight Loss That Fits, 2022. Accessed: 2022-04-11.

[105] YouAte. https://youate.com/, 2022. Accessed: 2022-04-11.

[106] Gabrielle M Turner-McGrievy, Michael W Beets, Justin B Moore, Andrew T Kaczynski, Daheia J Barr-Anderson, and Deborah F Tate. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mhealth weight loss program. *Journal of the American Medical Informatics Association*, 20(3):513–518, 2013.

[107] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, pages 3207–3216, 2015.

[108] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6850–6863, 2017.

[109] Elizabeth V Eikey and Madhu C Reddy. " it's definitely been a journey" a qualitative study on how women with eating disorders use weight loss apps. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 642–654, 2017.

[110] Elizabeth Victoria Eikey, Madhu C Reddy, Kayla M Booth, Lynette Kvasny, Johnna L Blair, Victor Li, and Erika S Poole. Desire to be underweight: Exploratory study on a weight loss app community and user perceptions of the impact on disordered eating behaviors. *JMIR mHealth and uHealth*, 5(10), 2017.

[111] Ferran Altarriba Bertran, Samvid Jhaveri, Rosa Lutz, Katherine Isbister, and Danielle Wilde. Making sense of human-food interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[112] Rohit Ashok Khot, Florian Mueller, et al. Human-food interaction. *Foundations and Trends® in Human-Computer Interaction*, 12(4):238–415, 2019.

[113] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. When personal tracking becomes social: Examining the use of instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on human factors in computing systems*, pages 1674–1687, 2017.

[114] Nađa Terzimehić, Christina Schneegass, and Heinrich Hussmann. Towards finding windows of opportunity for ubiquitous healthy eating interventions. In *International Conference on Persuasive Technology*, pages 99–112. Springer, 2018.

[115] Andrea Grimes and Richard Harper. Celebratory technology: new directions for food research in hci. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 467–476, 2008.

[116] G Bird and Peter Creighton Elwood. The dietary intakes of subjects estimated from photographs compared with a weighed record. *Human nutrition. Applied nutrition*, 37(6):470–473, 1983.

[117] Ian Li, Anind K Dey, and Jodi Forlizzi. Understanding my data, myself: supporting self-reflection with ubicomp technologies. *Ubicomp 2011*, pages 405–414.

[118] Brian Y Lim, Xinni Chng, and Shengdong Zhao. Trade-off between automation and accuracy in mobile photo recognition food logging. In *Proceedings of the Fifth International Symposium of Chinese CHI*, pages 53–59, 2017.

[119] Norman Azah Anir, Hairul Nizam, and Azmi Masliyana. The users perceptions and opportunities in malaysia in introducing rfid system for halal food tracking. *WSEAS Transactions on information science and applications*, 5(5):843–852, 2008.

[120] Jisu Jung, Lyndal Wellard-Cole, Colin Cai, Irena Koprinska, Kalina Yacef, Margaret Allman-Farinelli, and Judy Kay. Foundations for systematic evaluation and benchmarking of a mobile food logger in a large-scale nutrition study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–25, 2020.

[121] Hyungik Oh, Jonathan Nguyen, Soundarya Soundararajan, and Ramesh Jain. Multimodal food journaling. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, pages 39–47, 2018.

[122] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 451–462, 2016.

[123] Joan Burton. Who healthy workplace framework and model. *Geneva, Switzerland: World Health Organisation*, Chapter 3: What is a Healthy Workplace?, 2010.

[124] Gary White, Zilu Liang, and Siobhan Clarke. A quantified-self framework for exploring and enhancing personal productivity. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2019.

[125] Alexis Hiniker, Sungsoo Hong, Tadayoshi Kohno, and Julie A Kientz. Mytime: Designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 4746–4757, 2016.

[126] Andre N Meyer, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. Retrospecting on work and productivity: A study on self-monitoring software developers' work. *PACMHCI*, 1(CSCW):79–1, 2017.

[127] Prithwiraj Choudhury, Cirrus Foroughi, and Barbara Larson. Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal*, 42(4):655–683, 2021.

[128] SeekTech. Forest - stay focused, be present. https://www.forestapp.cc/, 2021. Accessed: 2022-04-11.

[129] RescueTime. RescueTime. https://www.rescuetime.com, 2008. Accessed: 2022-04-11.

[130] TimeCamp. TimeCamp. https://www.timecamp.com/, 2021. Accessed: 2022-04-11.

[131] Young-Ho Kim, Eun Kyoung Choe, Bongshin Lee, and Jinwook Seo. Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[132] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla NY Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. Large scale analysis of multitasking behavior during remote meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.

[133] Gerard A Kennedy and Harry Ball. Power break: A brief hypnorelaxation program to reduce work-related fatigue and improve work satisfaction, productivity, and well-being. *Australian Journal of Clinical and Experimental Hypnosis*, 35(2):169–193, 2007.

[134] Yuhan Luo, Bongshin Lee, Donghee Yvette Wohn, Amanda L Rebar, David E Conroy, and Eun Kyoung Choe. Time for break: Understanding information workers' sedentary behavior through a break prompting system. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[135] Scott A Cambo, Daniel Avrahami, and Matthew L Lee. Breaksense: Combining physiological and location sensing to promote mobility during work-breaks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3595–3607, 2017.

[136] Dan Morris, AJ Brush, and Brian R Meyers. Superbreak: using interactivity to enhance ergonomic typing breaks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pages 1817–1826. ACM, 2008.

[137] Health & fitness apps continue to trend upwards in 2021. https://www.adjust.com/blog/health-fitness-apps-trend-upwards-in-2021/. Accessed: 2022-04-11.

[138] Leib Litman, Zohn Rosen, David Spierer, Sarah Weinberger-Litman, Akiva Goldschein, and Jonathan Robinson. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. *Journal of medical Internet research*, 17(8):e195, 2015.

[139] Jennifer K Carroll, Anne Moorhead, Raymond Bond, William G LeBlanc, Robert J Petrella, and Kevin Fiscella. Who uses mobile phone health apps and does use matter? a secondary data analytics approach. *Journal of medical Internet research*, 19(4):e125, 2017.

[140] Ron Shor and Anat Shalev. Barriers to involvement in physical activities of persons with mental illness. *Health promotion international*, 31(1):116–123, 2014.

[141] Eveliina E Korkiakangas, Maija A Alahuhta, and Jaana H Laitinen. Barriers to regular exercise among adults at high risk or diagnosed with type 2 diabetes: a systematic review. *Health promotion international*, 24(4):416–427, 2009.

[142] Cindy HP Sit, John H Kerr, and Irene TF Wong. Motives for and barriers to physical activity participation in middle-aged chinese women. *Psychology of Sport and exercise*, 9(3):266–283, 2008.

[143] Centers for Disease Control and Prevention: Physical Activity Basics. https://www.cdc.gov/physicalactivity/basics/index.htm. Accessed: 2022-04-11.

[144] Elizabeth Jane Lyons and Claire Hatkevich. Prevalence of behavior changing strategies in fitness video games: theory-based content analysis. *Journal of medical Internet research*, 15(5):e81, 2013.

[145] David E Conroy, Chih-Hsiang Yang, and Jaclyn P Maher. Behavior change techniques in top-ranked mobile apps for physical activity. *American journal of preventive medicine*, 46(6):649–652, 2014.

[146] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 457–466. ACM, 2006.

[147] Andrew Prestwich, Macro Perugini, and Robert Hurling. Can the effects of implementation intentions on exercise be enhanced using text messages? *Psychology and Health*, 24(6):677–687, 2009.

[148] Nike Training Club App. https://www.nike.com/us/en_us/c/nike-plus/training-app. Accessed: 2022-04-11.

[149] SWORKIT. https://sworkit.com/. Accessed: 2022-04-11.

[150] Leap Fitness. 30 Day Fitness Challenge - Workout at Home. https://play.google.com/store/apps/details?id=com.popularapp.thirtydayfitnesschallenge. Accessed: 2022-04-11.

[151] Keep Workout and Fitness Trainer. https://www.keepkeep.com/. Accessed: 2022-04-11.

[152] Thomas Stütz, Michael Domhardt, Gerlinde Emsenhuber, Daniela Huber, Martin Tiefengrabner, Nicholas Matis, and Simon Ginzinger. An interactive 3d health app with multimodal information representation for frozen shoulder. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 3. ACM, 2017.

[153] Thomas Stütz, Gerlinde Emsenhuber, Daniela Huber, Michael Domhardt, Martin Tiefengrabner, Gertie Janneke Oostingh, Ulrike Fötschl, Nicholas Matis, and Simon Ginzinger. Mobile phone–supported physiotherapy for frozen shoulder: Feasibility assessment based on a usability study. *JMIR rehabilitation and assistive technologies*, 4(2):e6, 2017.

[154] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):70, 2018.

[155] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. Caring for vincent: A chatbot for self-compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[156] Lark. https://www.lark.com/. Accessed: 2022-04-11.

[157] HealthBot. https://healthbot.in/. Accessed: 2022-04-11.

[158] Upkar Varshney. Mobile health: Four emerging themes of research. *Decision Support Systems*, 66:20–35, 2014.

[159] Google Home. https://store.google.com/gb/product/google_home, 2022. Accessed: 2022-04-11.

[160] Amazon. Marketplace: Alexa skills. https://www.amazon.com/Marketplace, 2021. Accessed: 2022-04-11.

[161] Explore Google Assistant. https://assistant.google.com/explore/, 2022. Accessed: 2022-04-11.

[162] The 30 best Alexa skills in 2022. https://www.tomsguide.com/round-up/best-alexa-skills, 2022. Accessed: 2022-04-11.

[163] 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):e174, 2018.

[164] Alexa Skills Marketplace - Health & Fitness. https://tinyurl.com/2y6nxmjx, 2022. Accessed: 2022-04-11.

[165] 7 Min Fitness Challenge. http://7minworkoutapp.net/, 2022. Accessed: 2022-04-11.

[166] 5-Minute Plank Workout. https://www.amazon.com/cubic-ai-5-Minute-Plank-Workout/dp/B06XHTCB3Z. Accessed: 2022-04-11.

[167] Fitbit Skill. https://www.amazon.com/dp/B01CH4BP28, 2022. Accessed: 2022-04-11.

[168] My fitnesspal Amazon Alexa Skill. https://www.amazon.com/gp/product/B07QN179C5, 2022. Accessed: 2022-04-11.

[169] Markku Turunen, Jaakko Hakulinen, Olov Ståhl, Björn Gambäck, Preben Hansen, Mari C Rodríguez Gancedo, Raúl Santos de La Cámara, Cameron Smith, Daniel Charlton, and Marc Cavazza. Multimodal and mobile conversational health and fitness companions. *Computer Speech & Language*, 25(2):192–209, 2011.

[170] Ugan Yasavur, Christine Lisetti, and Naphtali Rishe. Let's talk! speaking virtual counselor offers you a brief intervention. *Journal on Multimodal User Interfaces*, 8(4):381–398, 2014.

[171] Tim Brown et al. Design thinking. *Harvard business review*, 86(6):84, 2008.

[172] Roger Martin and Roger L Martin. *The design of business: Why design thinking is the next competitive advantage*. Harvard Business Press, 2009.

[173] Chadia Abras, Diane Maloney-Krichmar, and Jenny Preece. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

[174] Elizabeth B-N Sanders and Pieter Jan Stappers. Co-creation and the new landscapes of design. *Co-design*, 4(1):5–18, 2008.

[175] Elizabeth B-N Sanders, Eva Brandt, and Thomas Binder. A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th biennial participatory design conference*, pages 195–198. ACM, 2010.

[176] Toni Robertson and Jesper Simonsen. Participatory design. *Routledge international handbook of participatory design*, page 1, 2012.

[177] Matthew K Hong, Udaya Lakshmi, Thomas A Olson, and Lauren Wilcox. Visual odls: Co-designing patient-generated observations of daily living to support data-driven conversations in pediatric care. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*, page 476. ACM, 2018.

[178] Michael J Gonzales and Laurel D Riek. Co-designing patient-centered health communication tools for cancer care. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pages 208–215. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013.

[179] Yoojung Kim, Eunyoung Heo, Hyunjeong Lee, Sookyoung Ji, Jueun Choi, Jeong-Whun Kim, Joongseek Lee, and Sooyoung Yoo. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*, pages 5787–5799. ACM, 2017.

[180] Greg Walsh, Alison Druin, Mona Leigh Guha, Elizabeth Foss, Evan Golub, Leshell Hatley, Elizabeth Bonsignore, and Sonia Franckel. Layered elaboration: a new technique for co-design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1237–1240. ACM, 2010.

[181] Marilyn Mcgee-Lennon, Aidan Smeaton, and Stephen Brewster. Designing home care reminder systems: lessons learned through co-design with older users. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 49–56. IEEE, 2012.

[182] Jonathan Lazar, Caitlin Woglom, Jeanhee Chung, Alison Schwartz, Yichuan Grace Hsieh, Richard Moore, Drew Crowley, and Brian Skotko. Co-design process of a smart phone app to help people with down syndrome manage their nutritional habits. *Journal of usability studies*, 13(2):73–93, 2018.

[183] Craig M MacDonald and Michael E Atwood. Changing perspectives on evaluation in hci: past, present, and future. In *CHI'13 extended abstracts on human factors in computing systems*, pages 1969–1978. 2013.

[184] Mary Beth Rosson and John M Carroll. *Usability engineering: scenario-based development of human-computer interaction*. Morgan Kaufmann, 2002.

[185] Thomas T Hewett, Ronald Baecker, Stuart Card, Tom Carey, Jean Gasen, Marilyn Mantei, Gary Perlman, Gary Strong, and William Verplank. *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992.

[186] Lucas M Silva, Yuqi Huai, and Daniel A Epstein. Exploring voice input opportunities in multimodal food journaling. In *CHI'20 Workshop on Conversational Agents for Health and Wellbeing*. 2020.

[187] Nanna Gorm and Irina Shklovski. Participant driven photo elicitation for understanding activity tracking: Benefits and limitations. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1350–1361, 2017.

[188] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. Activity tracking in vivo. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[189] Andrea Grimes, Martin Bednar, Jay David Bolter, and Rebecca E Grinter. EatWell: sharing nutrition-related memories in a low-income community. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 87–96, 2008.

[190] Verónica Rivera-Pelayo, Angela Fessl, Lars Müller, and Viktoria Pammer. Introducing mood self-tracking at work: Empirical insights from call centers. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(1):1–28, 2017.

[191] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. Reflect, not regret: Understanding regretful smartphone use with app feature-level analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–36, 2021.

[192] Gillian O'Loughlin, Sarah Jane Cullen, Adrian McGoldrick, Siobhan O'Connor, Richard Blain, Shane O'Malley, and Giles D Warrington. Using a wearable camera to increase the accuracy of dietary analysis. *American journal of preventive medicine*, 44(3):297–301, 2013.

[193] Meredith A Case, Holland A Burwick, Kevin G Volpp, and Mitesh S Patel. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *Jama*, 313(6):625–626, 2015.

[194] James Codella, Chohreh Partovian, H-Y Chang, and C-H Chen. Data quality challenges for person-generated health and wellness data. *IBM Journal of Research and Development*, 62(1):3–1, 2018.

[195] Wael Khazen, Jean-François Jeanne, Laëtitia Demaretz, Florent Schäfer, Guy Fagherazzi, et al. Rethinking the use of mobile apps for dietary assessment in medical research. *Journal of medical Internet research*, 22(6):e15619, 2020.

[196] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. Im2calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1233–1241, 2015.

[197] Nutritionix Nutrition API. https://www.nutritionix.com/business/api, 2022. Accessed: 2022-04-11.

[198] Edamam Nutrition Analysis API. https://developer.edamam.com/edamam-nutrition-api, 2022. Accessed: 2022-04-11.

[199] Eun Kyoung Choe, Bongshin Lee, Sean Munson, Wanda Pratt, and Julie A Kientz. Persuasive performance feedback: The effect of framing on self-efficacy. In *AMIA annual symposium proceedings*, volume 2013, page 825. American Medical Informatics Association, 2013.

[200] Antonio Bascur, Pedro O Rossel, Valeria Herskovic, and Claudia Martínez-Carrasco. Evitapp: persuasive application for physical activity and smoking cessation. In *Multidisciplinary Digital Publishing Institute Proceedings*, volume 2, page 1208, 2018.

[201] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 2653–2662, 2015.

[202] Amid Ayobi, Paul Marshall, Anna L Cox, and Yunan Chen. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6889–6901, 2017.

[203] Bernd Ploderer, Wally Smith, Steve Howard, Jon Pearce, and Ron Borland. Things you don't want to know about yourself: ambivalence about tracking and sharing personal information for behaviour change. In *Proceedings of the 24th australian computer-human interaction conference*, pages 489–492, 2012.

[204] Lora E Burke, Melanie Warziski, Terry Starrett, Jina Choo, Edvin Music, Susan Sereika, Susan Stark, and Mary Ann Sevick. Self-monitoring dietary intake: current and future practices. *Journal of Renal Nutrition*, 15(3):281–290, 2005.

[205] Lora E Burke, Molly B Conroy, Susan M Sereika, Okan U Elci, Mindi A Styn, Sushama D Acharya, Mary A Sevick, Linda J Ewing, and Karen Glanz. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. *Obesity*, 19(2):338–344, 2011.

[206] Kate EC Grimshaw, Joe Maskell, Erin M Oliver, Ruth CG Morris, Keith D Foote, EN Clare Mills, Barrie M Margetts, and Graham Roberts. Diet and food allergy development during infancy: birth cohort study findings using prospective food diary data. *Journal of Allergy and Clinical Immunology*, 133(2):511–519, 2014.

[207] Lashbrook Angela. The Next Wellness Trend Should Be Google Spreadsheets. https://onezero.medium.com/the-next-wellness-trend-should-be-google-spreadsheets-4675ae6a8496#79ba. Accessed: 2022-04-11.

[208] Klahre Ayn-Monique. How To Use a Bullet Journal to Meal Plan and Track Your Food Budget, According to the Creator. https://www.thekitchn.com/bullet-journal-food-budget-meal-plan-253166. Accessed: 2022-04-11.

[209] John Pruitt and Tamara Adlin. *The persona lifecycle: keeping people in mind throughout product design*. Elsevier, 2010.

[210] Cynthia LeRouge, Jiao Ma, Sweta Sneha, and Kristin Tolle. User profiles and personas in the design and development of consumer health technologies. *International journal of medical informatics*, 82(11):e251–e268, 2013.

[211] Carolyn Snyder. *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann, 2003.

[212] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2015.

[213] Sketch. https://www.sketchapp.com/. Accessed: 2022-04-11.

[214] Recovery Record. Recovery Record – Best Practice for Eating Disorders Treatment, 2022. Accessed: 2022-04-11.

[215] David S Ludwig and Jon Kabat-Zinn. Mindfulness in medicine. *Jama*, 300(11):1350–1352, 2008.

[216] Eun Kyoung Choe, Bongshin Lee, Tariq Osman Andersen, Lauren Wilcox, and Geraldine Fitzpatrick. Harnessing the power of patient-generated data. *IEEE Pervasive Computing*, 17(2):50–56, 2018.

[217] Mollie McKillop, Lena Mamykina, and Noémie Elhadad. Designing in the dark: Eliciting self-tracking dimensions for understanding enigmatic disease. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*, page 565. ACM, 2018.

[218] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. Co-designing food trackers with dietitians: Identifying design opportunities for food tracker customization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[219] Jeffery Sobal and Carole A Bisogni. Constructing food choice decisions. *Annals of Behavioral Medicine*, 38(suppl_1):s37–s46, 2009.

[220] Gastón Ares, Franco Mawad, Ana Giménez, and Alejandro Maiche. Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference*, 31:28–37, 2014.

[221] Deborah A Cohen and Susan H Babey. Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity Reviews*, 13(9):766–779, 2012.

[222] Benjamin Schüz, Natalie Schüz, and Stuart G Ferguson. It's the power of food: individual differences in food cue responsiveness and snacking in everyday life. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):149, 2015.

[223] Marcela CC Bomfim, Sharon I Kirkpatrick, Lennart E Nacke, and James R Wallace. Food literacy while shopping: Motivating informed food purchasing behaviour with a situated gameful app. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[224] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3988–3999, 2016.

[225] Young-Ho Kim, Bongshin Lee, Jinwook Seo, and Eun Kyoung Choe. OmniTrack for Research: A Research Platform for Streamlining Mobile-based In Situ Data Collection, 2019.

[226] Els Bilman, Ellen van Kleef, and Hans van Trijp. External cues challenging the internal appetite control system—overview and practical implications. *Critical reviews in food science and nutrition*, 57(13):2825–2834, 2017.

[227] Samsung, Inc. Samsung Voice Recorder, 2022. Accessed: 2022-04-11.

[228] Mohd Khanapi Abd Ghani and Ika Novita Dewi. Comparing speech recognition and text writing in recording patient health records. In *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, pages 365–370. IEEE, 2012.

[229] Zoom Video Communications, Inc. Video Conferencing, Web Conferencing, Webinars, Screen Sharing - Zoom, 2022. Accessed: 2022-04-11.

[230] TeamViewer. TeamViewer QuickSupport. https://www.teamviewer.com/en-us/info/quicksupport/, 2022. Accessed: 2022-04-11.

[231] Yunlong Wang, Ulrike Pfeil, and Harald Reiterer. Supporting self-assembly: The ikea effect on mobile health persuasive technology. In *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*, pages 19–22, 2016.

[232] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[233] Jolene D Smyth, Don A Dillman, Leah Melani Christian, and Mallory McBride. Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 73(2):325–337, 2009.

[234] WA Van Staveren, JO De Boer, and J Burema. Validity and reproducibility of a dietary history method estimating the usual food intake during one month. *The American journal of clinical nutrition*, 42(3):554–559, 1985.

[235] Lisa B Signorello, Heather M Munro, Maciej S Buchowski, David G Schlundt, Sarah S Cohen, Margaret K Hargreaves, and William J Blot. Estimating nutrient intake from a food frequency questionnaire: incorporating the elements of race and geographic region. *American journal of epidemiology*, 170(1):104–111, 2009.

[236] Margaret Raber, Tom Baranowski, Karla Crawford, Shreela V Sharma, Vanessa Schick, Christine Markham, Wenyan Jia, Mingui Sun, Emily Steinman, and Joya Chandra. The healthy cooking index: Nutrition optimizing home food preparation practices across multiple data collection methods. *Journal of the Academy of Nutrition and Dietetics*, 2020.

[237] Arnel B Ocay, Jane M Fernandez, and Thelma D Palaoag. Nutritrack: Android-based food recognition app for nutrition awareness. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pages 2099–2104. IEEE, 2017.

[238] Mette Holst, Tina Beermann, Marie Nerup Mortensen, Lotte Boa Skadhauge, Marianne Køhler, Karen Lindorff-Larsen, and Henrik Højgaard Rasmussen. Optimizing protein and energy intake in hospitals by improving individualized meal serving, hosting and the eating environment. *Nutrition*, 34:14–20, 2017.

[239] Susan Albers. *Eating mindfully: How to end mindless eating and enjoy a balanced relationship with food*. New Harbinger Publications, 2012.

[240] Amy Leung Hui, Gustaaf Sevenhuysen, Dexter Harvey, and Elizabeth Salamon. Barriers and coping strategies of women with gestational diabetes to follow dietary advice. *Women and Birth*, 27(4):292–297, 2014.

[241] 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.

[242] Eun Kyoung Choe, Bongshin Lee, and m.c. schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications*, 35(4):28–37, 2015.

[243] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia van Gemert-Pijnen. Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing*, 18(7):1667–1676, 2014.

[244] 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.

[245] 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.

[246] Rowanne Fleck and Geraldine Fitzpatrick. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, pages 216–223, 2010.

[247] Shuo Niu, D Scott McCrickard, Timothy L Stelter, Alan Dix, and G Don Taylor. Reorganize your blogs: Supporting blog re-visitation with natural language

processing and visualization. *Multimodal Technologies and Interaction*, 3(4):66, 2019.

[248] Daniel PW Ellis and Keansub Lee. Minimal-impact audio-based personal archives. In *Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 39–47, 2004.

[249] 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 *Proceedings of the 2021 Conference on Designing Interactive System*. ACM, 2021.

[250] Holly Seirup and Sage Rose. Exploring the effects of hope on gpa and retention among college undergraduate students on academic probation. *Education Research International*, 2011, 2011.

[251] Denise Pfeiffer. *Academic and environmental stress among undergraduate and graduate college students: A literature review*. PhD thesis, 2001.

[252] Peter Bates and Lori Goff. The invisible student: Benefits and challenges of part-time doctoral studies. *Alberta Journal of Educational Research*, 58(3):368–380, 2012.

[253] Christine Mollen. Trying to fit in: Barriers to degree completion for part-time graduate students, 2016.

[254] Jessica M Nicklin, Emily J Meachon, and Laurel A McNall. Balancing work, school, and personal life among graduate students: A positive psychology approach. *Applied Research in Quality of Life*, 14(5):1265–1286, 2019.

[255] Rajeev Darolia. Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, 38:38–50, 2014.

[256] Wolfgang Klein. The basic variety (or: Couldn't natural languages be much simpler?). In *Looking at Language*, pages 403–465. De Gruyter Mouton, 2018.

[257] Philipp Kirschthaler, Martin Porcheron, and Joel E Fischer. What can i say? effects of discoverability in vuis on task performance and user experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–9, 2020.

[258] Arjun Srinivasan and John Stasko. How to ask what to say?: Strategies for evaluating natural language interfaces for data visualization. *IEEE Computer Graphics and Applications*, 40(4):96–103, 2020.

[259] Emmanuel Vincent, Shinji Watanabe, Aditya Arie Nugraha, Jon Barker, and Ricard Marxer. An analysis of environment, microphone and data simulation mismatches in robust speech recognition. *Computer Speech & Language*, 46:535–557, 2017.

[260] Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. How blocking distractions affects workplace focus and productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 928–934, 2017.

[261] Kotlin. https://kotlinlang.org/. Accessed: 2022-04-11.

[262] Microsoft Cognitive Service - Speech to Text. https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/. Accessed: 2022-04-11.

[263] Smile - Statistical Machine Intelligence and Learning Engine. https://haifengl.github.io/nlp.html. Accessed: 2022-04-11.

[264] Natty. http://natty.joestelmach.com/. Accessed: 2022-04-11.

[265] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[266] Virginia Braun and Victoria Clarke. Thematic analysis. 2012.

[267] Ciaran B Trace and Yan Zhang. Minding the gap: Creating meaning from missing and anomalous data. *Information & Culture*, 56(2):178–216, 2021.

[268] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77, 2003.

[269] Diego Reforgiato Recupero, Erik Cambria, and Emanuele Di Rosa. Semantic sentiment analysis challenge at eswc2017. In *Semantic Web Evaluation Challenge*, pages 109–123. Springer, 2017.

[270] Microsoft Cognitive Service - Prepare data for Custom Speech. https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-custom-speech-test-and-train. Accessed: 2022-04-11.

[271] Google Cloud - Send a recognition request with speech adaptation. https://cloud.google.com/speech-to-text/docs/context-strength. Accessed: 2022-04-11.

[272] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110, 2008.

[273] Rosemary Pimm, Corneel Vandelanotte, Ryan E Rhodes, Camille Short, Mitch J Duncan, and Amanda L Rebar. Cue consistency associated with physical activity automaticity and behavior. *Behavioral Medicine*, 42(4):248–253, 2016.

[274] Justin Yang, Costas A Christophi, Andrea Farioli, Dorothee M Baur, Steven Moffatt, Terrell W Zollinger, and Stefanos N Kales. Association between push-up exercise capacity and future cardiovascular events among active adult men. *JAMA network open*, 2(2):e188341–e188341, 2019.

[275] Emmanuel Stamatakis, I-Min Lee, Jason Bennie, Jonathan Freeston, Mark Hamer, Gary O'Donovan, Ding Ding, Adrian Bauman, and Yorgi Mavros. Does strength-promoting exercise confer unique health benefits? a pooled analysis of data on 11 population cohorts with all-cause, cancer, and cardiovascular mortality endpoints. *American journal of epidemiology*, 187(5):1102–1112, 2017.

[276] Phillippa Lally and Benjamin Gardner. Promoting habit formation. *Health Psychology Review*, 7(sup1):S137–S158, 2013.

[277] RunKeepper. https://runkeeper.com/running-app. Accessed: 2022-04-11.

[278] JEFIT. https://www.jefit.com/. Accessed: 2022-04-11.

[279] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2269–2278. ACM, 2014.

[280] Android Studio. https://developer.android.com/studio. Accessed: 2022-04-11.

[281] Alexa Skill Kit. https://developer.amazon.com/en-US/alexa/alexa-skills-kit. Accessed: 2022-04-11.

[282] Node.js. https://nodejs.org/en/. Accessed: 2022-04-11.

[283] Amazon Web Service (AWS) Lambda. https://aws.amazon.com/lambda/. Accessed: 2022-04-11.

[284] Google Firebase. https://firebase.google.com/. Accessed: 2022-04-11.

[285] Bas Verplanken and Sheina Orbell. Reflections on past behavior: a self-report index of habit strength. *Journal of Applied Social Psychology*, 33(6):1313–1330, 2003.

[286] Janice M Brown, William R Miller, and Lauren A Lawendowski. The self-regulation questionnaire. 1999.

[287] Harm op den Akker, Valerie M Jones, and Hermie J Hermens. Tailoring real-time physical activity coaching systems: a literature survey and model. *User modeling and user-adapted interaction*, 24(5):351–392, 2014.

[288] 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*, pages 5286–5297. ACM, 2016.

[289] Florence-Emilie Kinnafick, Cecilie Thøgersen-Ntoumani, and Joan L Duda. Physical activity adoption to adherence, lapse, and dropout: a self-determination theory perspective. *Qualitative health research*, 24(5):706–718, 2014.

[290] Lie Ming Tang and Judy Kay. Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):26, 2017.

[291] Jacob O Wobbrock and Julie A Kientz. Research contributions in human-computer interaction. *interactions*, 23(3):38–44, 2016.

[292] Lesley Strawderman and Rick Koubek. Human factors and usability in service quality measurement. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 18(4):454–463, 2008.

[293] Debjyoti Ghosh, Can Liu, Shengdong Zhao, and Kotaro Hara. Commanding and re-dictation: Developing eyes-free voice-based interaction for editing dictated text. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(4):1–31, 2020.

[294] Amazon Echo Show - Echo With Smart displays. https://www.amazon.com/echoshow, 2022. Accessed: 2022-04-11.

[295] Nassim Jafarinaimi, Jodi Forlizzi, Amy Hurst, and John Zimmerman. Breakaway: an ambient display designed to change human behavior. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1945–1948, 2005.

[296] Nicholas Epley, Adam Waytz, and John T Cacioppo. On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4):864, 2007.

[297] Wenhuan Wang. Smartphones as social actors? social dispositional factors in assessing anthropomorphism. *Computers in Human Behavior*, 68:334–344, 2017.

[298] Cassie A Ludwig, Natalia F Callaway, Joyce Ho Park, and Theodore Leng. Mobile health in the retinal clinic population: Access to and interest in self-tracking. *Ophthalmic Surgery, Lasers and Imaging Retina*, 47(3):252–257, 2016.