Data@Hand: Fostering Visual Exploration of Personal Data on Smartphones Leveraging Speech and Touch Interaction

Young-Ho Kim
University of Maryland
College Park, MD, USA
yghokim@umd.edu

Bongshin Lee
Microsoft Research
Redmond, WA, USA
bongshin@microsoft.com

Arjun Srinivasan*
Tableau Research
Seattle, WA, USA
arjunsrinivasan@tableau.com

Eun Kyoung Choe
University of Maryland
College Park, MD, USA
choe@umd.edu

Figure 1: Data@Hand supports multimodal interactions to enable people to easily navigate and compare their personal health data on smartphones. People can execute a context-agnostic command such as setting up a comparison by specifying two new periods using a global speech button ①. They can feed a context to their utterance by touch, such as the start date ②, the target for comparison ③, or the time range for refining the view ④. (Please refer to our supplementary video, available at `https://data-at-hand.github.io/chi2021`, which demonstrates the interactions.)

ABSTRACT

Most mobile health apps employ data visualization to help people view their health and activity data, but these apps provide limited support for visual data exploration. Furthermore, despite its huge potential benefits, mobile visualization research in the personal data context is sparse. This work aims to empower people to easily navigate and compare their personal health data on smartphones by enabling flexible time manipulation with speech. We designed and developed Data@Hand, a mobile app that leverages the synergy of two complementary modalities: speech and touch. Through an exploratory study with 13 long-term Fitbit users, we examined how multimodal interaction helps participants explore their own health data. Participants successfully adopted multimodal interaction (i.e., speech and touch) for convenient and fluid data exploration. Based on the quantitative and qualitative findings, we discuss design implications and opportunities with multimodal interaction for better supporting visual data exploration on mobile devices.

*Arjun Srinivasan conducted this work while with Georgia Institute of Technology.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445421

CCS CONCEPTS

• **Human-centered computing** → **Visualization**; **Visualization systems and tools**; **Empirical studies in visualization**; **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Personal informatics, data visualization, multimodal interaction, speech, smartphone

ACM Reference Format:

Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. 2021. Data@Hand: Fostering Visual Exploration of Personal Data on Smartphones Leveraging Speech and Touch Interaction. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.3445421

1 INTRODUCTION

Smartphones, equipped with high-resolution displays and powerful processors, are increasingly becoming a dominant way to access information [74]. A vast number of mobile health (or mHealth) apps, including wearable devices' companion apps (e.g., Fitbit App [30], Apple Health [3], Samsung Health [59], Garmin [33], and Mi Fit [36]), enable people to access their health and activity data collected over time. While these mHealth apps commonly employ data visualizations to help people view and understand personal data [2], they provide limited support for navigating and exploring the data. Furthermore, research on mobile data visualization is sparse [48]. Much of the mobile visualization research has been carried out with tablets [10], and only a handful of projects have recently begun to study data visualizations on smartphones (e.g., [11, 12, 61]) and smartwatches (e.g., [9, 13, 46]).

In this work, we investigate how to facilitate flexible data exploration on smartphones in the context of self-tracking data, while addressing several challenges smartphones pose. Due to their limited screen space, smartphones cannot afford a control panel of widgets (alongside the visualizations), which are effective means to support dynamic queries [1]. It is distracting to navigate to a separate page to adjust the widgets and come back to the page with visualizations to see the effect. In addition, the lack of mouse input makes it difficult to perform two essential actions—(1) a precise selection and (2) details-on-demand using a mouse hover interaction—which are well supported in a desktop environment. Furthermore, while *time* is a primary dimension of self-tracking data, it is laborious to perform time-based interactions on smartphones, such as entering specific date, time, and ranges. As a result, most mHealth apps tend to limit time manipulations. For example, the Fitbit App restricts people to view data by predefined time segments, such as one week, one month, three months, and one year. Inspired by previous research advocating the benefits of multimodal interaction [15, 20, 49, 50, 75], we incorporate an additional input modality, *speech*, to overcome these challenges. Speech-based interaction takes little screen space. Speech is flexible to cover different ways that people specify date (e.g., "Last Thanksgiving" or "Lunar New Year's Day") and date ranges (e.g., "2017" indicating the range from January 1, 2017 to December 31, 2017), which people are already familiar with.

Combining two complementary modalities, speech and touch, we designed and developed Data@Hand (Figure 1), a mobile app that facilitates visual data exploration. As a first step, informed by prior work on personal insights [17, 18], we support navigation and temporal comparisons of personal health data, as well as data-driven queries. To understand how speech and touch interaction can help lay individuals explore their data, we conducted an exploratory study with 13 long-term Fitbit users using Data@Hand.

We observed that participants successfully adopted multimodal interaction, using both speech and touch interactions while finding personal insights. Participants reported that they made deliberate choices between the two input modalities for a more convenient and fluid data exploration. Flexible time expressions enabled by speech-based natural language interaction helped them freely navigate data in a specific time frame (e.g., "*Go to March 2020*"), quickly set up comparisons (e.g., "*Compare sleep ranges of winter and summer this year,*" illustrated in ① in Figure 1), and easily execute data-driven queries (e.g., "*Days I walked more than 10,000 steps last month*"). Speech commands combined with touch input (e.g., ②, ③, and ④ in Figure 1) enabled easy modifications of the time components. For example, to change the start/end date, one can simply utter a specific date while holding on the start/end date label. Also, graphical widgets (e.g., calendar widget, data source drop-down list) served as a fallback to correct erroneous results of speech or to explore a set of categorical values. Being satisfied with their overall experiences, all but one participant expressed that they are willing to keep using Data@Hand after the study. The key contributions of this work are:

(1) The design and implementation of Data@Hand, the first mobile app that leverages the synergy of speech and touch input modalities for personal data exploration. Data@Hand helps people interact with their own personal data on smartphones by accessing the Fitbit data using the Fitbit REST API. It runs on both iOS and Android, using the Apple speech framework [4] and Microsoft Cognitive Speech API [53] as speech recognizers. The Data@Hand source code is available at `https://data-at-hand.github.io`.

(2) An empirical study conducted with 13 long-term Fitbit users using Data@Hand. From the quantitative and qualitative analysis, we provide an understanding of how people explore their own data using speech and touch interaction on smartphones, uncovering situations and rationale for people's choice of interaction.

(3) Design implications and opportunities for multimodal interaction for mobile data visualization. Reflecting on our observations and participants' feedback, we draw design implications and opportunities for developing a multimodal interaction to better support personal data exploration on mobile devices.

2 RELATED WORK

In this section, we cover related work in the areas of (1) visual exploration of personal data and (2) natural language and multimodal interaction for visual data exploration.

2.1 Visual Exploration of Personal Data

As collecting and reflecting on personal data has become commonplace, research on personal visualization has gained increasing attention [37, 49, 51, 73]. Personal visualizations equipped with interactivity enable *visual data exploration*, making it easy for people to understand and reflect on their data. As such, many Personal Informatics systems (e.g., [6, 18, 26, 28, 38, 72]) support visual data exploration to empower people to gain personal insights.

Because *time* is a primary dimension of personal data (or self-tracking data), systems that support visual data exploration strive to enable easy manipulation of the time component. For example, Visualized Self [18]—a web application that enables people to integrate and explore personal data from multiple self-tracking services—employs the timeline mini-map to enable rapid adjustment of the data scope. Activity River [6] takes a similar timeline-based approach, but the scope was fixed as a single day. Visual Mementos [72] supports visual exploration of personal location history. It incorporates a multidimensional selection widget for the precise scoping of event episodes in a series of location logs. Huang and colleagues [38] designed an on-calendar visualization tool that integrated people's physical activity data. They chose to leverage a calendar, an inherent time-based visualization with rich personal context, to make the data readily accessible. We note that these systems were designed for a desktop environment and did not investigate how their interfaces could be applicable in a mobile environment with smaller screen space.

Commercial mHealth apps, including wearable devices' companion apps, also provide the visual exploration capability. However, most of these apps are limited in terms of time navigation, making it hard to jump to an arbitrary time frame or to compare data from two different time frames. These commercial apps usually show daily information using a dashboard on their main page, aiming to promote self-awareness of the current performance. They are also constrained by the smartphone form factor, such as small screen and imprecise touch input. Furthermore, existing widgets (e.g., calendar) for date entry are not flexible enough to handle the various ways to specify time. As a result, limited navigation support becomes a barrier to performing flexible data exploration, and in turn obtaining personal data insights on smartphones.

The practitioner community developed ample applications of data visualization in mobile apps & websites (refer to [56, 57] for curated practices). Data visualization is also commonly used for mobile form factors, such as smartphones and tablets, in research prototypes (e.g., [16, 22, 41, 42, 55, 60]) developed by UbiComp and Human-Computer Interaction researchers. However, research specifically focusing on mobile data visualization is sparse and much of the mobile visualization research has been carried out with tablets [10]. As such, the research community has recently put efforts to shape a research agenda for mobile data visualization while calling for more research endeavors [14, 48, 49]. Although only a handful, mobile visualization research has begun to pay attention to the smaller form factors (i.e., smartphones and smartwatches). They examined effectiveness of visual representations (e.g., ranges on timeline [11], animated transition vs. small multiples of scatterplots [12], data comparison on smartwatches [9]) and interaction techniques (e.g., multivariate network exploration [25], pan and zoom timelines and sliders [61]). In addition, in their workshop paper, Choe and colleagues [15] envisioned a scenario where speech interaction could facilitate personal data exploration on mobile devices. Inspired by this line of research and vision, we contribute to mobile data visualization with Data@Hand, the first mobile app that leverages the synergy of speech and touch input modalities to augment personal data exploration.

2.2 Natural Language and Multimodal Interaction for Visual Data Exploration

Advancements in natural language understanding and speech recognition technology have promoted the design of natural language interfaces (NLIs) for data visualization [8, 21, 23, 32, 34, 35, 39, 40, 47, 62–65, 67, 69, 79]. These systems commonly exploit two advantages of natural language: (1) high flexibility in synthesizing multiple commands and parameters [21] and (2) low barriers in expressing intents and questions regarding the data [7, 66]. A majority of these systems are primarily designed for desktop settings and investigate typed natural language input (e.g., [21, 23, 32, 35, 62, 69]). On the other hand, another subset of prior systems focus on speech input and explore multimodal visualization interfaces that incorporate speech with other modalities such as pen and/or touch in post-WIMP [76] settings including tablets [39, 64] and large displays [47, 65, 67].

Given the context of the smartphone form factor, two existing systems that are most relevant to our work are the tablet-based systems, Valletto [39] and InChorus [64]. Valletto allows people to specify charts through speech and then perform simple touch gestures such as a rotate to flip axes and swipe to change the visualization type. Exploiting the complementary nature of pen/touch and speech [19], InChorus illustrates a vocabulary of multimodal actions involving a wide range of visualizations for tabular data (e.g., selecting marks with a pen and saying "*Remove others*" to filter unselected points, pointing on an axis with a finger and speaking attribute names to specify data mappings).

Our work extends this line of research on multimodal visualization interfaces on mobile devices in two notable ways. First, unlike Valletto and InChorus that were designed for tablets, Data@Hand is the first system specifically designed for smartphones. Correspondingly, we discuss the unique constraints we faced in the form of more limited screen space and lower precision with touch input [78], and how these constraints impacted our interface and interaction design (see DR3 in Section 3.1). Second, compared to prior NLIs that primarily target avid users of visualization tools (e.g., data analysts, developers or managers in visualization-oriented products), our work targets a broader population of lay individuals interested in exploring their personal health data. In doing so, we highlight the role of lay individuals that affected the design of Data@Hand, and discuss in detail the interaction patterns individuals employed for performing tasks such as time manipulation, temporal comparisons, and data-driven queries with Data@Hand.

3 DATA@HAND

To enable flexible data exploration on smartphones, we designed and developed Data@Hand, a mobile app that leverages the benefits of two complementary input modalities, speech and touch. In this section, we describe our design rationales and the Data@Hand app along with the implementation details.

3.1 Design Rationales

DR1: Use Simple and Familiar Visualizations to Support Lay Individuals. Our target audience is people who collect their health and activity data using commercial wearable devices. They usually do not have expertise in data visualization and analytics. Therefore, we incorporated familiar visual representations and conventional charts that people commonly encounter on existing mHealth apps. For example, we used bar charts, line charts, and range charts to visualize daily measurement values (e.g., step counts and resting heart rates in Figure 1a and sleep ranges in Figure 2a). As for visualizing the aggregated data over a period, we designed custom representations called *aggregation plots*: a simplified version of box-and-whisker plots. We encoded only the average and the minimum and the maximum values (for the range) because these metrics are more relevant to the personal tracking context than median and percentiles. For example, Figure 2c shows the aggregated sleep ranges for February and August of 2020.

DR2: Enable Flexible Time Manipulation to Help Identify Personal Insights. Unlike the general and broader visual data exploration, personal visualizations require different design considerations because of the nature of the personal data and the diverse personal data collection goals [37, 73]. In visual exploration with personal data (or self-tracking data), people look for specific personal insights, such as whether they achieved a certain personal goal (e.g., 10,000 steps per day), how their behaviors (e.g., step counts, sleep patterns) and emotional/physiological states (e.g., mood, heart rates) change over time, and what factors might have affected the changes (e.g., before & after the COVID-19 lockdown) [18]. Furthermore, as shown in prior works [17, 18, 51, 68], *comparison by time segmentation* is one of the most common visual exploration tasks that people actively perform to gain personal insights. However, flexible time manipulation—a key facilitator in drawing personal insights—is rarely supported in mHealth apps due to the limitations we described earlier (see Section 2.1). We strove to enable flexible time manipulation, focusing on navigation (Figure 1a) and temporal comparisons (Figure 1b & 1c and Figure 2c & 2d).

DR3: Leverage the Synergy of Speech and Touch Interactions on Smartphones. Smartphones are increasingly becoming a dominant way to access information [74], and much of the personal data is collected from smartphones and wearable devices. As such, we wanted to facilitate easy access to people’s own data on smartphones. To overcome the challenges smartphones pose, we leverage both speech and touch input modalities that are complementary in nature: speech input affords a high *freedom of expression* without requiring much screen space, whereas touch input supports *direct* interaction [58, 64, 65, 67]. When combining the strengths of these two modalities, we aimed to provide a complementary set of operations [20] rather than providing an equivalent set of operations for each modality. We detail how we synergistically incorporated the two modalities in supporting a diverse set of operations on smartphones in Section 3.2.2.

3.2 User Interface and Interaction Design

Data@Hand currently supports five health metrics that are retrieved from the Fitbit data sources: (1) step count, (2) resting heart rate, (3) sleep duration, (4) hours slept, and (5) weight. For sleep, we included only one sleep log per day that is marked as the main sleep. We chose the metrics based on their prevalence in commercial health apps according to a recent survey [44].

3.2.1 User Interface and Interaction Components. Data@Hand supports navigation, temporal comparison, and data-driven queries using four main pages—*Home* (Figure 1a), *Data Source Detail* (Figure 2a), *Two-range Comparison* (Figure 1b), and *Cyclical Comparison* (Figure 1c). As a default view, the Home page visualizes the

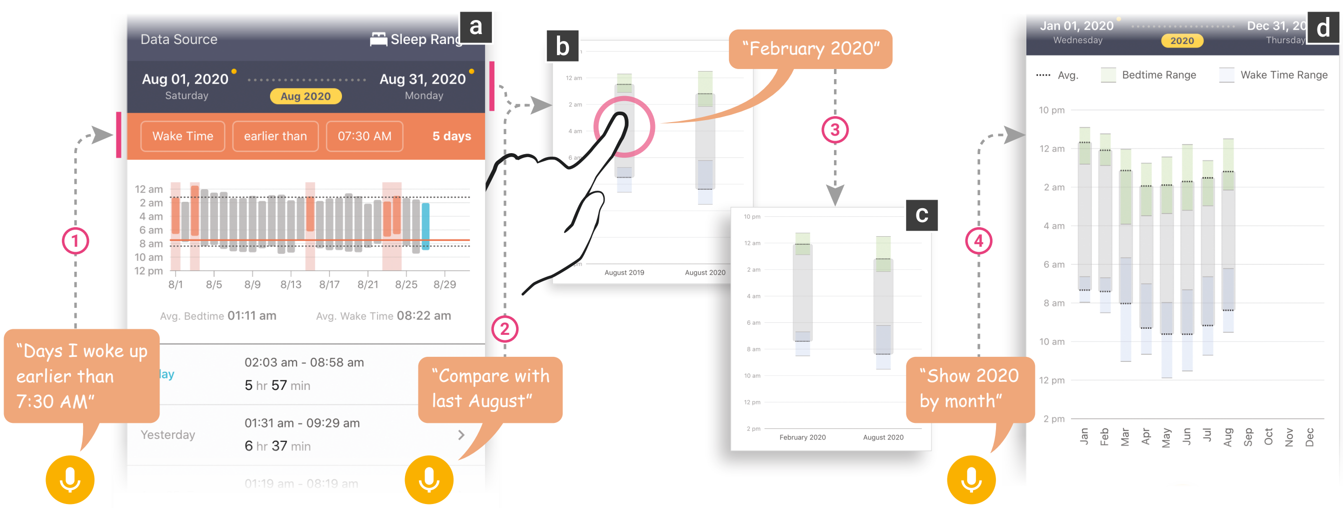

Figure 2: Excerpt of the exploration flow in our usage scenario. ① Zoe executes a data-driven query via natural language. ② The system infers omitted parameters (e.g., a pre-selected date range as a comparison target) using the current screen information. ③ Zoe changes August 2019 to February 2020 by touch+speech interaction on the aggregation plot. ④ Zoe establishes the cyclical comparison to see this year’s monthly trend.

past 7-day data for the five data sources. The Data Source Detail page shows detailed information for a single data source. Both comparison pages juxtapose aggregated measurement values, but in two different ways: the Two-range Comparison page plots two selected periods side by side, and the Cyclical Comparison page displays values within a specific period grouped by a predefined time cycle, such as days of the week (i.e., Sunday through Saturday) and months of the year (i.e., January through December).

These pages contain common interaction components. An app header and bottom toolbar are located on all pages (see Figure 1). The *range widget* on the app header is for manipulating date ranges, and the global microphone button on the toolbar is used to execute a speech command in a global scope—a command that is agnostic to the current context or page. We describe its functionalities in more detail in Section 3.2.2. The Home button is a shortcut that brings back to the Home page, maintaining the current date range. The Compare button opens the configuration panel where people can configure data source, comparison type, and date ranges to execute comparison queries.

People can execute *data-driven queries* by specifying a condition in natural language. Data@Hand responds by highlighting the data items in red (see Figure 2a). Also, the *query bar* (① in Figure 2a), shown below the app header, contains *parameter widgets* for manipulating recognized parameters in a query (e.g., [Wake Time], [earlier than], [07:30 AM]), and the number of days that satisfy the condition (e.g., '5 days'). The system automatically updates the query result when people manipulate the time or data source on the screen, until they dismiss the query bar by swiping it to the left. The design was inspired by the *ambiguity widgets* from prior systems (e.g., [32, 62, 65, 67]).

3.2.2 Speech and Touch Interaction. Using touch and speech input modalities, Data@Hand provides three types of interaction—*touch-only*, *speech-only*, and *touch+speech*. With Data@Hand, people can directly interact with graphical widgets with touch, as they would normally do with any mobile apps. For example, they can tap on an item (e.g., label, chart) to select it, and swipe the range widget to shift the time frame back and forth. Using speech, people can issue powerful commands that are applicable to a global scope: for example, changing a data source and date range together with *"Show me the step counts from this summer,"* or executing a data-driven query with "*Highlight the days I walked more than 10,000 steps*." To handle speech, Data@Hand adopts a "push-to-talk" technique: the system records the speech input while people are pressing on the global microphone button . Given that pressing on the global microphone button serves only as a means to initiate speech interaction, we consider using speech with the global microphone button as speech-only interaction.

For the touch+speech interaction, people press on a target element while uttering speech commands that make use of the specific context the target element provides. For example, as shown in ② in Figure 1, people can navigate to a different date range by uttering only a date (e.g., "*January 1*") while pressing on the "start date" part of the range widget. Furthermore, with touch+speech people can perform a related command while keeping the same context. For example, as shown in ③ in Figure 1, after comparing the sleep ranges of winter and summer of this year (2020), they can simply utter "*Summer 2019*" while pressing on the aggregation plot for the winter to compare summer of 2019 and 2020. As demonstrated in previous research [52, 64, 77], this context helps Data@Hand facilitate faster interactions and reduce the complexity of speech commands, and ultimately improves user experiences.

3.2.3 Speech Affordance and Feedback. When people press the global microphone button or long-press a target element for touch+speech interaction, Data@Hand displays the *speech input panel* (Figure 3) while dimming outside. If the target element is an aggregation plot, the speech input panel is embedded in a tooltip (Figure 3, left). As a preemptive guide, Data@Hand displays a contextual message (e.g., "Say a new date" for start/end date) or example phrases on the panel (Figure 3, left) until people start to utter. While people are talking to the system, the panel displays a real-time dictation result (Figure 3, right) to make people aware of how their utterance is being recognized and to prevent the early release of the finger before the utterance is completely dictated.

After each execution of an operation, Data@Hand provides different types of feedback depending on its result. When Data@Hand could translate the utterance and execute a valid operation, it momentarily displays a confirmation message (① in Figure 4) along with the undo button (② in Figure 4) as a quick recovery option. If the translated operation is invalid, the system opens a contextual message dialog (③ in Figure 4). For example, if one utters just one date range using speech-only interaction on the Two-range Comparison page, the system suggests try the same command using touch+speech interaction through the aggregation plot for disambiguation. When Data@Hand fails to translate the utterance, it informs people accordingly (④ in Figure 4).

3.3 Interacting with Data@Hand

Here we describe Data@Hand's interactions through a usage scenario: a self-tracker Zoe has been collecting data using a Fitbit band for almost five years. While this scenario emphasizes the speech-only and touch+speech interactions, most operations are also supported by touch-only interaction using graphical widgets. (Refer to our supplementary video to see more detailed interaction.)

Data Navigation & Data-Driven Queries. Being curious about her long-term activity patterns, Zoe opens Data@Hand. The system

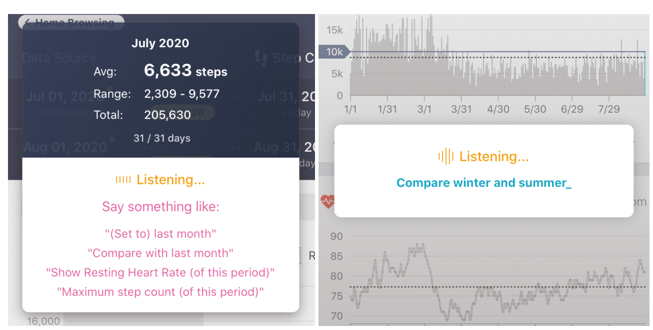

Figure 3: The speech input panel that displays preemptive guides (left) and a real time dictation result (right), before and during a speech interaction.

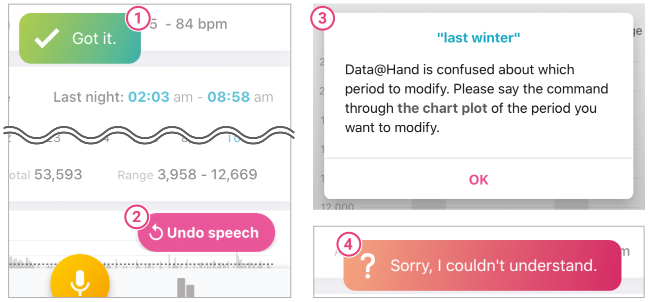

Figure 4: When the system could translate the utterance and execute a valid operation, it momentarily displays a confirmation message ① along with the undo button ②. When the translated operation is not valid, the system opens a contextual message dialog ③. When the system fails to translate the utterance, it informs people accordingly ④.

initially shows the past seven days of data on the Home page. To extend the scope, she long-presses the start date and utters "*January 1*" (② in Figure 1), and the system sets the date range to January 1, 2020 to today (August 27, 2020). Skimming the bar chart for step counts, she notices a dramatic drop of step counts since mid-March, being reminded of the start of the COVID-19 lockdown. This plummet motivates her to explore her data since the lockdown has taken effect.

Zoe starts to explore her step counts wanting to see how many days she achieved her daily step goal. She speaks "*Days I met my step goal.*" Referring to her step count goal (10,000 steps) from her Fitbit account, Data@Hand highlights the days with step counts higher than 10,000. She finds that the highlighted days are dense earlier in the year but sparse since March. Zoe scrolls through the charts for other data sources. Once she reaches the chart for sleep range, she notices that her sleep has been pushed back since March. Seeing this, Zoe decides to take a detailed look at her sleep ranges. To narrow down the scope to a more recent period, she speaks "*Sleep range of this month.*" The system opens the Data Source Detail page for August's sleep range. By asking "*Days I woke up earlier than 7:30 AM,*" she learns that she woke up that early just for five days in August (Figure 2a).

Temporal Comparisons. Zoe is curious about how her sleep differs from last year's. She utters "*Compare with last August*" (② in Figure 2). Translating last August to *August 2019*, Data@Hand opens the Two-range Comparison page comparing August 2019 against August 2020 (Figure 2b). Zoe notices that this August's average sleep schedule is shifted by more than an hour late compared to August 2019. Also, the ranges of the bedtime and wake time in August 2020 are longer, implying her irregular sleep pattern. Zoe now wonders how the lockdown has affected her sleep. So, she changes the range from *August 2019* to *February 2020*, the previous month before the lockdown, by uttering "*February 2020*" while long-pressing the aggregation plot for August 2019 (③ in Figure 2). She confirms that, compared to February, her sleep schedules for August are also shifted towards later hours and show more irregular bedtime and wake time (Figure 2c).

To see how the lockdown affected her sleep schedules from the monthly trend, she speaks "*Show 2020 by month*" (④ in Figure 2). Data@Hand opens the Cyclical Comparison page, with her sleep ranges in 2020 grouped by month (Figure 2d). She learns that her average sleep schedules have become more regular since May and that they have shifted to earlier hours. Zoe continues on the exploration, switching to other data sources by swiping the *data source widget* on the app header.

3.4 Implementation

We implemented Data@Hand in TypeScript [54] upon React Native [27] to support both iOS and Android. When a participant first signs in with a Fitbit account through OAuth 2, the system uses Fitbit REST API [31] to download the Fitbit data of the entire period since the account creation to the local SQLite database. The system always uses the locally-cached data to improve the performance by minimizing network overheads.

We used the Apple speech framework [5] and Microsoft Cognitive Services [53] as a speech-to-text recognizer on iOS and Android, respectively. We initially used a built-in speech recognizer for each OS. However, for Android, we decided to use Microsoft's Speech service because of the limitations of Android's built-in speech recognizer API. We appended a set of application-specific keywords (e.g., name of the data sources) and time expressions (e.g., "May" is likely to refer a month rather than a verb) to the recognizers' vocabulary to improve accuracy for short phrases.

We implemented the system interpreter to work locally on the device. Receiving the recognized input text, the interpreter preprocesses it by performing part-of-speech tagging using the Compromise [43] Javascript library and identifying parameters such as data sources, query conditions, and periods. To identify the time information mentioned in the input text, we used a customized version of Chrono [70], a natural language time parsing library. After the preprocessing, the interpreter infers the operation based on the tagged verbs and parameters, the current screen information, and the pressed element for the touch+speech interaction.

4 USER STUDY

We conducted an exploratory study with Data@Hand, employing a think-aloud protocol, to examine how multimodal interactions help people explore their data. As part of this study, participants interacted with their own Fitbit data using their smartphones. Due to the COVID-19 outbreak, all study sessions were held remotely using Zoom video call (in July 2020). In Section 4.2, we explain precautionary action we took to deliver a remote tutorial and to ensure close monitoring of the study session while mitigating potential privacy invasion. We refined both the system and study protocol through six pilot sessions with Fitbit users recruited from Reddit. This study was approved by the Institutional Review Board of the University of Maryland at College Park.

4.1 Participants

We recruited 13 participants (P1–13; nine females and four males) from Reddit by advertising the study on the subreddits for job postings in the United States. Our inclusion criteria were adults

Table 1: Summary of demographic and the Fitbit experience of our study participants.

Alias	Age	Gender	Occupation	Fitbit usage	Fitbit wearable	Collected data
P1	27	M	Client services trainer	2y 4m	Versa lite edition	Step, HR, Sleep, Weight (Aria)
P2	27	F	Chemical engineer	4y 6m	Alta	Step, HR, Sleep
P3	28	F	Freelance social media manager	4y 10m	Versa	Step, HR, Sleep
P4	32	M	Implementation specialist	5y 1m	Ionic	Step, HR, Sleep, Weight (BT)
P5	35	F	Freelance photographer	5y 2m	Blaze	Step, HR, Sleep, Weight (BT)
P6	23	F	Graduate student	4y 8m	Alta HR	Step, HR, Sleep
P7	23	F	Graduate student	5y 6m	Charge 2	Step, HR, Sleep, Weight
P8	46	F	Product manager	1y 7m	Versa	Step, HR, Sleep, Weight
P9	30	F	Healthcare consultant	5y 9m	Versa	Step, HR, Sleep, Weight (Aria)
P10	24	M	Unemployed (Varsity football player)	1y 8m	Versa	Step, HR, Sleep, Weight
P11	28	M	Software engineer	4y 1m	Charge 3	Step, HR, Sleep, Weight
P12	36	F	Professional figure skater	2y 1m	Alta	Step, HR, Sleep, Weight
P13	39	F	Market survey manager	4y 4m	Alta	Step, HR, Sleep, Weight

BT: Bluetooth scales which are from other vendors but can feed the data to Fitbit.

who (1) are native English speakers; (2) have used Fitbit wearables for at least six months and tracked at least three of the following measures: step count, heart rate, sleep, and weight; (3) are interested in looking at their Fitbit data; (4) are currently using iPhone or Android; (5) have no visual, motor, or speech impairments; (6) have used voice recognition systems within six months with a generally positive experience; and (7) can understand simple charts.

The demographic and Fitbit usage information of our study participants is shown in Table 1. Participants' ages ranged from 23 to 46 (avg = 30.62). Ten participants were full-time employees, two were graduate students, and one was unemployed. At the time of the study, participants had used Fitbit for an average of 47 months based on their account creation date. Four participants had been tracking weight using Aria [29] or third-party Bluetooth scales (which share the data with Fitbit). The screen sizes of participants' smartphones ranged from 4.7 to 6.1 inches (eight participants used iPhones with a 4.7-inch screen). We offered a 30 USD Amazon gift card for their participation.

4.2 Study Setup and Procedure

4.2.1 Pre-study Preparation. A fully-remote study using participants' own data required us to take extra precautions, such as mitigating potential invasion of privacy from the use of participants' own smartphones, preparing the training material, and establishing robust audio- and video- recording methods. Furthermore, because Fitbit allowed only 150 API calls per hour per account, we had to prefetch the Fitbit data before the study session. (Immediately after the study session, we deleted the participants' Fitbit data from our server.) To do so, we sent participants a link to a web page where they could sign an electronic consent form and fill out a pre-study questionnaire asking their Fitbit usage patterns and experiences of using voice assistant systems. After completing the questionnaire, the participants were asked to sign in with their Fitbit accounts so that our crawler could cumulatively download participants' entire Fitbit data. We also delivered the Data@Hand app to participants through TestFlight (iOS) or Google Play beta testing (Android) so that they could install the app on their phone.

4.2.2 Remote Study Session. Participants joined a 90-minute study session via Zoom video call [80] from their computer. Figure 5 illustrates the settings of the remote study session. Via TeamViewer QuickSupport [71], participants shared their smartphone screen with the experimenter. Prior to the screen sharing, the experimenter instructed participants to remove any privacy-sensitive information from their home screen and to turn off all the notifications. The experimenter then shared his monitor screen with the participant using Zoom's screen sharing feature: the participant could see how their smartphone screen was being displayed to the experimenter. Using the recording feature in Zoom, the experimenter recorded the video call session including the shared smartphone screen.

Tutorial. After explaining the goal of the study, the experimenter gave a 40-minute tutorial, using an example dataset containing fabricated data generated based on the first author's four years of Fitbit data. The tutorial covered Data@Hand's key design components and interactions—data sources & charts, time manipulation, data navigation, temporal comparison, and data-driven queries. The experimenter introduced and demonstrated each feature using presentation slides and a video clip on a shared screen (refer to our supplementary material available at `https://data-at-hand.github.io/chi2021`), and gave participants a chance to practice before moving to a new feature. In particular, participants were encouraged to practice the push-to-talk interaction exploring the example dataset. We gave them example speech queries (via shared Zoom screen) that they could use for practice, but also encouraged them to try out their own speech commands. We gave them enough time to practice until they feel comfortable using both the touch and speech input to interact with Data@Hand.

Free-form Exploration. In this phase, the experimenter instructed participants to freely explore their own data with Data@Hand. We asked them to use *any* of the supported interaction modalities

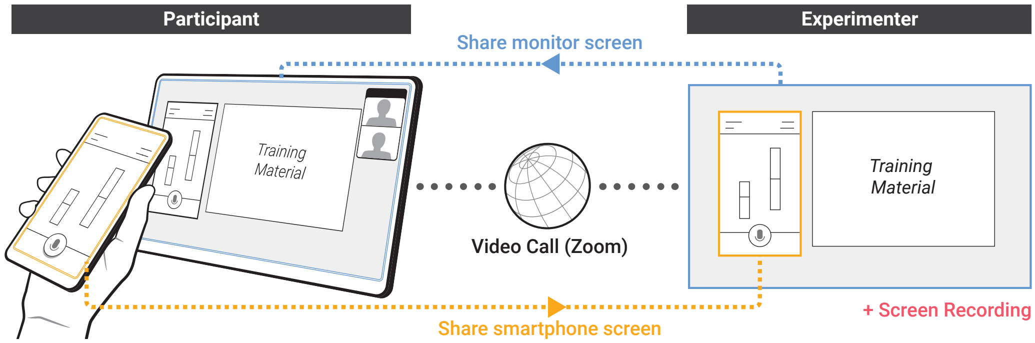

Figure 5: Settings of the remote study session using video call and screen sharing. The participant shares the smartphone screen with the experimenter, and the experimenter shares his monitor screen with the participant.

(touch, speech, touch+speech) of their choice. We also asked them to think aloud as they explored the data so that we can capture the insights they found and the challenges they faced, and understand their intentions and experiences. The experimenter observed how participants interacted with Data@Hand through the shared screen showing their smartphone screen and faces. We audio- and video-recorded each video call session and the system logged the interaction history and uploaded the logs to our server. This free-form exploration phase lasted approximately 20 minutes.

Debriefing. We conducted a semi-structured interview around 10–15 minutes at the end of the session. We asked participants about their experiences with Data@Hand, difficulties and confusing features, and follow-up questions based on our observation in the exploration phase. We also asked them about the use cases of Data@Hand they could envision and their willingness to keep using the Data@Hand app after the study.

4.3 Data Analysis

We analyzed the video recordings and the interaction logs from the free-form exploration phase. We performed both quantitative and qualitative analysis to examine how participants used the speech and touch modalities in finding personal insights. As for the quantitative analysis, we first extracted participants' interaction attempts to perform actions for data navigation, temporal comparisons, and data-driven queries, reviewing the exploration videos and interactions logs. We defined an *interaction attempt* as a series of low-level interactions (e.g., tapping, swiping) that were involved to obtain a desired outcome. To modify a start date, for example, people may tap the start date on the range widget to invoke a calendar popup and tap the target date. We treated this series of tap interactions as one attempt with touch-only interaction.

As for the qualitative analysis, we analyzed the transcripts from the exploration phase to identify the types of personal insights, following Choe and colleagues' definition of *personal insight* ("an individual observation about the data accompanied by visual evidence") [17, 18]. We extracted personal insights and categorized their types. For example, we extracted the following from P10's exploration session: (On the Cyclical Comparison page) "*Just pretty interesting that I get my most steps on Saturdays.*" We coded this observation with three insight types: *extreme* ("*most steps*"), *reference* ("*Saturdays*"), and *comparison by time segments* (essential to identify the day with the most step counts in this case). We describe when and how participants gained the insights in Section 5.2.

We transcribed the audio recordings of the debriefing interviews, which were grouped according to the following aspects: (1) participants' rationales of choosing the input modalities; (2) new analyses/tasks/questions Data@Hand enabled; (3) challenges participants encountered; and (4) how participants envisioned the use cases of Data@Hand in their own contexts. We referenced this information when interpreting the video recordings and interaction logs, as well as to understand participants' general reactions to Data@Hand (reported in Section 5.3).

5 RESULTS

In this section, we report the results of our study in three parts: (1) interaction usages, (2) personal insights, and (3) general reactions to Data@Hand.

5.1 Interaction Usages

We identified 809 interaction attempts in total. Among these, 400 (49.4%) were touch-only, 281 (34.7%) were speech-only, and 128 (15.8%) were touch+speech interaction attempts. Among the 400 touch-only attempts, five were aborted by participants to perform the equivalent action using speech instead (e.g., P4 first opened a calendar picker to modify the start date, but he closed it and modified the start date using the touch+speech interaction). Among the 281 speech-only attempts, 32 were failed due to the recognition/interpretation errors, 16 were invalid actions (e.g., "*Compare hours slept*" without designating any comparative periods), and five were unsupported actions (e.g., attempting to execute a data-driven query to the aggregation plots). Among the 128 touch+speech attempts, eight were recognition/interpretation errors, and seven were invalid actions (e.g., uttering a date where a period is required). As a result, 736 (395 touch-only, 228 speech-only, and 113 touch+speech) interaction attempts were successfully executed, which we call *operations* from now on. Of these, we included only 589 operations that are relevant to the three main features (time manipulation, temporal comparison, and data-driven queries) into further analysis (268 touch-only, 209 speech-only, and 112 touch+speech operations). The rest 147 operations consisted of 127 touch-only and 19 speech-only operations for data source manipulation, and one touch+speech operation for data-driven query. Figure 6 visualizes these operations by participant. In the following subsections, we describe participants' detailed usage of these input modalities for the three main features.

5.1.1 *Time Manipulation.*

In total, participants manipulated time 470 times (see Table 2 for the summary). Participants specified time (**T1–5**; e.g., change the start/end date) or manipulated time as part of executing comparison (**C1–4**) or data-driven queries (**Q1**). When manipulating time, participants actively used both speech and touch: to navigate to a new date range, participants used speech-only interaction 71 times (**T3**) and a calendar picker with touch 48 times (**T1**). To modify ranges in the Two-range Comparison page, participants tended to use touch+speech interaction on the aggregation plot instead of touch-only interaction. As shown in Table 3-top, 12 participants used touch+speech 37 times (**T5**) while 3 participants used touch-only 13 times (**T1**).

When participants modified only start or end date of the date range, their behaviors differed depending on the distance between their target date and the currently selected one. If the target date was close to the original one (especially within the same month), they preferred a calendar picker (**T1**) as it would require only a few taps. On the other hand, if the target date was far from the currently selected date (e.g., several months away), they preferred touch+speech interaction (**T4**) by long-pressing the date label and mentioning the target date.

Three participants (P5, P11, P13) heavily used swipe to modify the date range (**T2**, 146 out of 170 times). For example, starting from the Data Source Detail page for weight in year 2020, P13 swiped through the year 2016, skimming the trend of each year (see swiping sequences in Figure 6). The touch-only swipe is a quick way to navigate through using a preset date range.

5.1.2 *Temporal Comparisons.*

Table 3 summarizes the operations to execute comparisons (**C1–4**), including the cases that modify time as part of the follow-up comparisons (**T1–5**). As shown in Figure 6, participants often performed a series of comparisons (184 operations, units in both green and yellow without an **X** mark) by refining the time range (or the data source). When executing comparison queries (**C1–4**), participants tended to choose the input modalities depending on the type of comparison. For **two-range comparison**, participants were inclined to use speech-only interaction (**C2**): only three participants used the Compare button to execute the two-range comparison query with touch-only interaction (**C1**, 4 instances). On the other hand, participants showed mixed patterns on modality for **cyclical comparison**: among 12 participants who used the cyclical comparison, four participants (P1, P6, P12, and P13) used only touch (**C3**), two (P2 and P3) used only speech (**C4**), and the other six used both modalities.

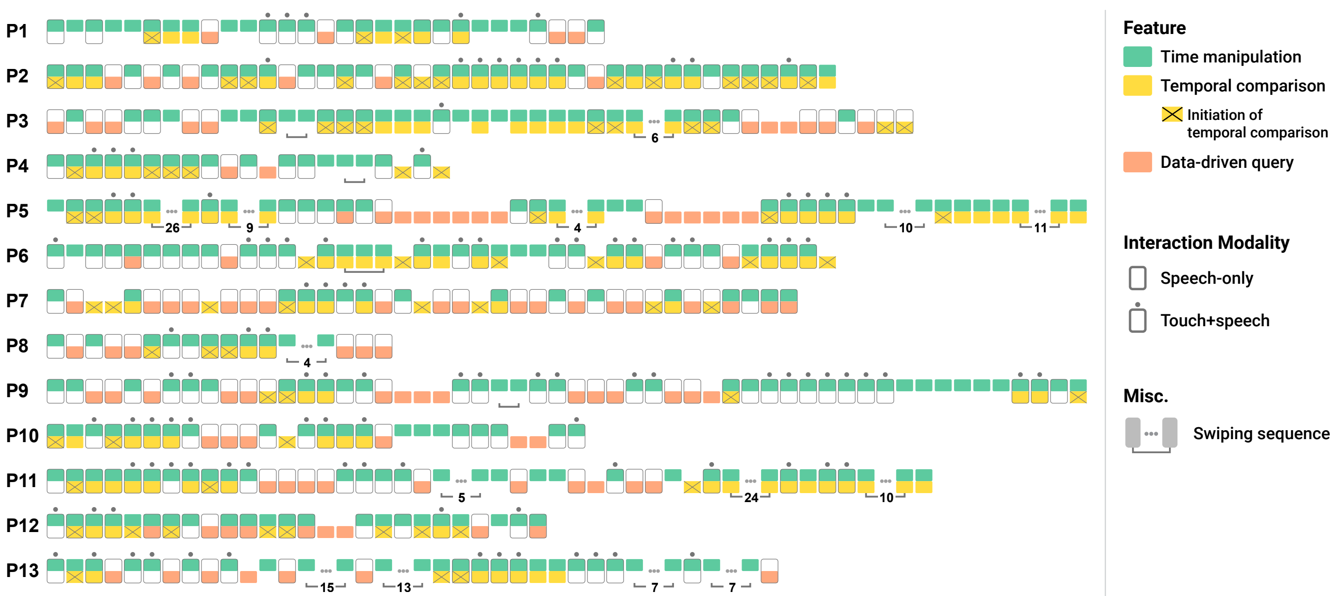

Figure 6: Sequences of operations—successfully executed interaction attempts—that are relevant to the three main features (time manipulation, temporal comparison, and data-driven query) by participant. Each unit on the horizontal-axis represents one operation, and the color of rectangles in a unit indicates the intended feature. The X mark indicates the initiation of temporal comparison (and thus only applicable for yellow rectangles; see C1–4 in Table 3). The border of a unit indicates the use of the speech modality (i.e., speech-only or touch+speech). A series of swiping to manipulate time is bundled or collapsed with a black horizontal line. This operation overview shows that participants used all three interaction types to perform various actions. Also, the prevalence of a series of the green+yellow units without X suggests that participants often performed a series of comparisons with time refinement.

Table 2: Summary of operations that contributed to the time manipulation, with the number of occurrences per participant and example utterances from participants. The modality column indicates the input modality of the operation (T: touch-only, S: speech-only, TS: touch+speech). T1–5 indicate the operations to directly manipulate time, whereas C1–4 and Q1 indicate the operations to execute the comparison or data-driven queries with time parameters.

Action	Operation Pattern	Modality	Total	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
Modify time directly	**T1.** Tap start/end date label then pick date	T	48	11	1	13	1	6	3			3	3	4	1	2
	T2. Swipe range widget until reaching the target	T	170			8	2	62	4		4	5	1	41		43
	T3. Speak <(data source) period > "*Last 30 days*" / "*Step count in 2019*"	S	71	6	11	5	6	5	7	8	3	7	6	4	3	
	T4. start/end date label and speak < date > start date + "*January 1, 2019*"	T S	75	4		1	1	1	14		1	18	8	13	4	10
	T5. aggregation plot and speak < period > In the two-range comparison page, the left period plot + "*January 2020*"	T S	37	1	10		3	6	3	4	2	2	1	1	1	3
[a]**Execute comparison query**	**C1+3.** Tap Compare button then configure parameters	T	13	1		1		1	2			1			5	2
	C2. Speak <Compare (date source) period (period)> "*Compare January 2018 with January 2019*"	S	30	2	4	6	3	3		1	3	2	1	1	3	1
	C4. Speak <(data Source) cycle period > "*Show me sleep by month for 2020*"	S	16		8	3	1	1				1	1	1		
[b]**Execute data-driven query**	**Q1.** Speak < condition period > "*Maximum step count last month*"	S	10					1	1	3					5	
(): There exist operations which do not include this parameter.			470	25	34	37	17	86	34	16	13	39	21	65	22	61

[a]The occurrences are a subset of the operations of the equivalent patterns in Table 3.
[b]The occurrences are a subset of the operations of the equivalent patterns in Table 4.

5.1.3 Data-Driven Queries. Table 4 summarizes the operations dedicated to initializing and editing a data-driven query. Data@Hand supported only speech-only interaction to initialize a query (**Q1**), with touch-only interaction only for follow-up editing of the recognized parameters through the widgets on the query bar (**Q2**). In total, participants executed data-driven queries 89 times (avg = 6.85). Eight participants edited a recognized query parameter using a parameter widget 24 times. A majority of the data-driven queries were invoked to identify extreme values (e.g., lowest step count) or days when they achieved a goal (e.g., days with step counts higher than 10,000). However, participants also used data-driven queries to identify unusual days (e.g., days with step counts less than 3,000 steps for lazy days [P5], days with bedtimes later than 5:00 AM for the days with sleep troubles [P12]).

5.2 Personal Insights

We extracted 367 data-driven observations (avg = 28.23, min = 10, max = 52) from 13 participants and derived 694 personal insights. Table 5 summarizes the insight type categories, frequency, and example quotes for each category (refer to [18] for definition of each category). Overall, participants gained various types of insights, covering most of those observed with a desktop personal data exploration tool [18] and from the data presentation videos of quantified-selfers (the enthusiastic self-trackers) [17], except three categories: *comparison by multiple services*, *prediction*, and *distribution by category*. In this section, we highlight notable categories and how Data@Hand supported gaining the insights.

Participants found 143 instances of **comparison** insights leveraging Data@Hand's two types of comparison: two-range comparison and cyclical comparison. They actively drew existing knowledge, or **external contexts**, which were not captured in the data, in executing comparison queries; these contexts often served as a **factor** of comparison. For example, most participants were interested in comparing their activity level before and after the stay-at-home orders around mid-March 2020 caused by the COVID-19 outbreak; common patterns for this were to compare a recent month (after the lockdown) with the same month of previous years (e.g., July 2019 vs. July 2020), and to investigate the monthly **trend** of year 2020. Other factors participants considered include job changes, start of a new project, and school semesters. When participants were curious about a past period, they usually compared it with the recent period (e.g., "*This month*," "*Last 90 days*") as a reference.

Participants compared not only values (e.g., measurement values or their average) but also other aspects of their data, such as **trends** and **variability or consistency**. In the comparison page, participants often inferred the variability from the aggregation plots, as each plot showed the range of the values (e.g., "[Looking at the days-of-the-week comparison screen] ...*the time I went to bed was most inconsistent on Mondays, Wednesdays, and*

Table 3: Summary of operations that contributed to establish a new comparison. C1–4 indicate the operations to execute a comparison query and T1–5 indicate the operations to directly manipulate time during the comparison, with their occurrences divided by the type of comparison.

Action	Operation Pattern	Modality	Total	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
Two-range Comparison																
Execute two-range comparison query	C1. Tap Compare button then configure period (period)	T	4			1			2						1	
	C2. Speak <Compare (date source) period (period)>	S	30	2	4	6	3	3		1	3	2	1	1	3	1
[a] Modify time directly	T1. Tap start/end date label then pick date	T	13	3		8										2
	T2. Swipe range widget until reaching the target	T	45			6		39								
	T4. start/end date label and speak <date>	T S	5					1				2			2	
	T5. aggregation plot and speak <period>	T S	37	1	10		3	6	3	4	2	2	1	1	1	3
Cyclical Comparison																
Execute cyclical comparison query	C3. Tap Compare button then configure (data source) cycle (period)	T	22	1			2	1	4	5		1	1	1	4	2
	C4. Speak <(data Source) cycle (period)>	S	19		9	3	1	1		2		1	1	1		
[b] Modify time directly	T1. Tap start/end date label then pick date	T	3		1			2								
	T2. Swipe range widget until reaching the target	T	54					13	4				1	36		
	T3. Speak <period>	S	10	1	3					3			1	2		
	T4. start/end date label and speak <date>	T S	17						5				4	7		1
(): There exist operations which do not include this parameter.			259	8	27	24	9	66	18	15	5	8	10	49	11	9

[a, b] The occurrences are a subset of the operations of the equivalent patterns in Table 2.

Table 4: Summary of operations to execute data-driven queries. Q1 indicates the operations to execute a query in natural language, and Q2 indicates the operations to use the parameter widgets on the query bar to edit the recognized query.

Operation Pattern	Modality	Total	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
Q1. Speak <condition (period)> *"Days I slept more than six hours"*	S	89	4	6	9	1	3	4	19	6	11	4	9	7	6
Q2. Tap parameter widget then modify query parameter	T	24			2	1	11				4	2	1	2	1

(): There exist operations which do not include this parameter.

was most inconsistent on Mondays, Wednesdays, and Saturdays." – P7). About a half of the variability instances (23 out of 49) were inferred from the aggregation plots. When comparing *trends* among different periods, however, participants used working memory by sequentially navigating to each period because the comparison screens only provided with the aggregated information.

Eight participants sought **correlation** insights by associating values from different data sources on the same day. The highlights provided in response to data-driven queries served as visual links, as the same days were highlighted across all data sources. Guided by data-driven queries, participants often scrolled through the charts on the Home page or switched the data source on the Data Source Detail page, to identify similar patterns of peaks and drops against the superposed highlights. For instance, P13 highlighted days with low resting heart rate (lower than 56) on the heart rate detail page, and then navigated to the step count page, finding the similarity between the highlighted days and the daily step count values on those days ("*...the days my heart rate was low, was the days with my step count also, fairly low, for most days*" – P13).

5.3 General Reactions to Data@Hand

In the debriefing interview, we gathered participants' feedback on Data@Hand. In general, participants expressed excitement about the flexible time navigation and comparison features Data@Hand offered. Participants described time manipulation in natural language to be *fast* and *flexible*. For instance, P7 remarked,

> *I liked that I was able to change the date using the speech because I thought that was really easy rather than having to go through all of the different dates and months. And also it was cool to say like 'around this date' or 'this month' and it would get what you were talking about, whereas I think it could be hard to do in a traditional touch format.*

Nine participants contrasted their experience with Data@Hand to their previous experience with the Fitbit App, which was tedious

Table 5: Visualization insight types identified from the transcripts from the exploration phase, frequency with the number of participants (Ps), and example quotes. The insight types are sorted by frequency.

Insight Type and Frequency		Example Quotes
Detail (263)	**Identify value** (109) - 12Ps	"*What was that day? I got so much sleep. Wow, I got 13 and a half hours on that Monday.*" – P10
	Identify references (104) - 13Ps	"*Um.. wonder why it's slow in April.*" – P2 "*It's pretty consistent, but in December I had a very broad range the hours.*" – P6
	Identify extreme (50) - 12Ps	"*It looks like December was probably my highest activity month if I look back at the whole trend.*" – P8
Comparison (143)	**Of two instances** (66) - 11Ps	"*It's interesting to see my step count average is I was much more all over the place in October. But in February, I was much more consistent, which is better.*" – P1
	By time segments (43) - 10Ps	"*My average was cut about a half from January, February, down to March.*" – P4
	By factor (33) - 10Ps	"*My average (step count) was obviously higher six months ago than it is now because we're all locked in our houses.*" – P8
	Against external data (1) - 1P	"*I've been tracking what I eat. So I definitely used to weigh more. Look at that. this is like 150 lb.*" – P11
Recall (115)	**External context** (79) - 13Ps	"*That[hourly step count chart] would remind me of what I did that day. I know, based on the time of day and the day of the week, that it was a hike that I went on and that's why I got the extra steps.*" – P8
	Confirmation (26) - 8Ps	"*I know I've been getting less sleep recently compared to before. That's what I wanted to see.*" – P11
	Contradiction (10) - 5Ps	"*...because nothing feels the same (after the COVID-19 outbreak). But it's interesting to see that the data looks less terrible than I expected, and I'm kind of happy.*" – P12
Value judgment (51) - 10Ps		"*My average bedtime is.. somewhere between 2 and 4 am..pretty terrible.*" – P9
Variability (49) - 12Ps		"*Looks like I mean for everything in 2020, it seems to be my sleep is getting much more consistent, which is a good sign.*" – P1
Trend (42) - 11Ps		"*It's an increasing trend but like it's super low in the beginning of April which I find odd.*" – P2
Correlation (18) - 8Ps		"*The days that I have done the least amount of steps are the days of my heart rate is the lowest on average. That makes sense.*" – P3
Outlier (7) - 3Ps		"*Wow, there is definitely an outlier there.*" – P1
Data summary (6) - 3Ps		"*My average steps is 10760. Wow 4 millions (of total steps) *laugh* that's a lot. Range anywhere from eight to twenty four thousand, because sometimes I didn't wear it.*" – P6

to reach a specific date or period. P1 mentioned, "*If I was in the Fitbit App, I would have to go through a whole bunch of different screens to find one specific date and all of that data.*" Participants also appreciated being able to view two periods side-by-side in the Two-range Comparison page. Seven participants reported that they wanted similar feature in the Fitbit App, with which they currently have to sequentially view the two periods, while relying on working memory. P3 remarked, "*It [Fitbit App] just isn't as efficient, but I definitely have gone and looked back. But it took so much time to do it, but by the time I got to the date from last year and then I got back to today, I forgot [the last year's value].*"

Lastly, to gauge the utility of Data@Hand beyond the study session, we asked participants if they would be willing to use Data@Hand in real life (Data@Hand on their phones will continue to work by retrieving recent data directly from the Fitbit server.) All but one participant said that they would keep the Data@Hand app for continued use after the study. One participant decided not to keep the app because she had not been engaging in physical activities as usual due to the COVID-19 lockdown.

6 DISCUSSION

In this section, we discuss lessons learned from the design and implementation of Data@Hand as well as from the exploratory study. We also reflect on implications for better supporting personal data exploration through multimodal interaction in mobile contexts.

6.1 Enhancing Personal Data Exploration through Flexible Time Manipulation

Enabling flexible time manipulation was one of our key design rationales (DR2), which we strongly believe facilitates the insights gaining process. We observed that Data@Hand's time manipulation feature was well received by participants, mainly due to the speech that enabled flexible expression of time that participants are already familiar with. To support time-based data exploration with personal data, a system needs a time range specified by start and end time points (this is commonly applied in typical graphical range widgets). People, however, express time in a variety of ways, not necessarily specifying two time points. Examples include the range using the present date as an implicit anchor (e.g., "*Last six months*"), and the range presented with a semantic phrase (e.g., "*This March*" indicating the range from March 1, 2020 to March 31, 2020). Furthermore, people may not remember the date for notable events (e.g., the summer solstice) and some events are not represented with a fixed date (e.g., Thanksgiving in the United States). Accommodating these diverse time expressions in Data@Hand enabled participants to easily and flexibly set time ranges.

Major speech recognizers handle national holidays and seasons ingrained in our culture, but they do not know key personal events (e.g., job change, vacation, semester), which are important in exploring personal data. Thus, our study participants had to rely on their memory to perform time-based exploration using their personal contexts. Tagging personally meaningful events and being able to refer to them with speech could address this issue, offering great opportunities to enhance personal data exploration (e.g., "*Compare my sleep range between spring semester and summer break*").

Although our work focused on a much-needed mobile scenario of exploring personal health data, we envision that multimodal interactions leveraging time expressions can be similarly useful in other contexts. For example, flexible time navigation and comparison features can facilitate exploring personal data beyond the health and fitness domains, such as productivity, finance, and location history (e.g., "*Compare last week's screen time with this week's*," "*Most expensive expense during the last quarter*," "*Show me the places that I visited in the last three months*").

6.2 Complementary Roles of Speech and Touch Input Modalities

Our quantitative results showed that participants used both speech and touch modalities, individually and in tandem, performing all three types of interactions—touch-only, speech-only, and touch+speech. Our observations and participants' feedback also suggest that participants made deliberate choices between the two input modalities. In the debriefing interview, participants distinguished the advantages of the two modalities.

Speech interaction was generally considered to be fast and flexible, especially when making big shifts in terms of time ranges, or executing commands involving multiple keywords (e.g., "*Compare step counts of this month and last month,*" from the Home page). P13 remarked, "*I would like to be able to use voice for more things just because it's easier and I've found myself just sometimes thinking like, 'Oh, would this be quicker if I use a voice command?' 'Would it be easier if I use a voice command?'*" On the other hand, **touch** interaction was preferred in some cases, such as shifting time frames successively with swipe (see the swiping sequences from Figure 6) or choosing a data source from a list with a tap. This is also reflected in the high number of touch-only operation for data source manipulation (127 out of 146). In addition, participants resorted to touch when they were having a difficulty remembering exact keywords for a speech command (e.g., "month-of-the-year" for the cyclical comparison). Participants favored the **touch+speech** interaction when refining pre-executed commands (e.g., uttering a new date while pressing on the date button to modify only the start date). P4 noted, "*I felt much more confident doing that [touch+speech] because I knew that it was only got to manipulate that one aspect of a comparison chart or only the start date, instead of having to be more precise with my speech in what I was asking.*"

6.3 Natural Language vs. Keyword-Based Commands

We observed two different patterns of participants' use of speech commands—natural language (e.g., the two-range comparison and data-driven queries) and keyword-based utterances (e.g., uttering "*Hours slept*" to set the data source). They impose different technical challenges. The natural language commands were sensitive to the linguistic structure of participants' utterance. All 18 interpretation errors (the system recognized the utterance correctly but did not interpret the recognized text successfully) occurred during the natural phrasing of commands. To prevent such errors and improve the interpretation coverage of the system, we can collect utterance examples via crowdsourcing or from pilot studies to identify common linguistic structures people use to perform interactions.

On the other hand, keyword-based commands were vulnerable to recognition errors [45] with the generalized speech-to-text recognizers we used. All eight errors related to the keyword phrases occurred at the recognizer level. Such recognition errors may be prevented by training the recognizer with keywords as shown in recent projects [64, 65]. However, using a customized recognizer is currently not feasible for smartphone apps without involving a remote server, which may cause additional delays and thus hamper the user experience.

6.4 Designing Multimodal Interaction for Smartphones

Our main design goal with Data@Hand was to support the visual exploration of self-tracking data on smartphones. The smartphone form factor and personal data context led to the design choices that are different from general-purpose multimodal data exploration systems on tablet devices, such as InChorus [64] and Valletto [39].

Multimodal interaction of InChorus and Valletto focuses on constructing visualization (e.g., performing data binding and visual encoding, specifying chart types) and their exploration is driven by attributes in a given table. For example, InChorus incorporates a wide range of multimodal interaction to support a flexible visualization construction on tablets: people can choose the input modality they prefer, such as drag-and-drop, point & tap (using two fingers), point and write (with a pen) or speak, utter a command with speech, to perform a mapping between data attributes and chart elements, such as axes and legends. Valletto supports simple touch gestures such as swiping and rotating for manipulating visual encoding (e.g., flipping the axes) and provides a persistent chat panel where people can speak to generate a new chart, exclude/include attributes, or ask analytic questions such as correlation between two attributes. However, it would be challenging to provide all of such interactions on a smartphone, which has a smaller screen, needs to be held with a non-dominant hand, and may not support a pen.

Furthermore, in the personal data context especially to assist lay individuals, it is not necessary to support flexible visualization construction: it is more important to facilitate easy navigation and comparison across the time dimension in a given chart (e.g., bar chart and line chart). Therefore, Data@Hand's multimodal interaction focuses on the manipulation of time parameters, while reducing the complexity of interface.

Our study results suggest that participants can learn and use such multimodal interaction to find various insights from their self-tracking data, and their reactions were generally positive. We believe that Data@Hand's multimodal interaction achieved a good balance between flexibility and learnability by carefully considering smartphone form factor and personal data context together. In addition, we note that, for the personal data exploration, Data@Hand's interaction can be transferred to tablet form factors.

6.5 Privacy Concerns on Voice Interaction for Personal Data

While describing their real-world use cases in the debriefing, seven participants noted that they would be inclined to use only touch in the public space for two main reasons: (1) they did not want to disturb others and (2) they were afraid that surrounding people might feel awkward seeing them verbalizing health-related queries. For example, P3 mentioned, "*If you're only able to use speech, there's no privacy. You can't be at the doctor's office and be like, 'Tell me how much weight I've lost' or 'What day was I the fattest last year.'*" These remarks align with the findings from previous research that privacy concerns can discourage the use of voice interaction in close proximity to other people [24], thus potentially limiting the applicability of speech-incorporated multimodal interaction in the public setting.

6.6 Improving Visual Representations

According to our design rationale (DR1), we used basic charts that many people are already familiar with, such as bar charts and line charts. We also designed a new representation (i.e., aggregation plots) to support temporal comparison, which requires data aggregations. Furthermore, to efficiently communicate results for the data-driven queries, we enhanced basic charts with a highlighting capability while presenting data without aggregation (i.e., one bar/dot per day) in the Home page.

We see opportunities to improve visual representations to further enhance data exploration experiences especially with long-term data. While participants could view year-long data without aggregation (which was shown to be readable in recent research [11]), the current charts would not scale to view a longer term beyond a year. One straightforward solution would be to use our aggregation plots with data grouping (e.g., by week, by month, by year), a common approach in existing mobile apps. However, the lower level of details induced by the aggregation makes it difficult to highlight particular data points. It is an open research question to effectively show query results on these aggregation plots, for example, highlighting days with steps over 10,000 on the aggregation plots grouped by month.

6.7 Conducting an Exploratory Study in a Remote Environment

We had to address a number of challenges to convert the original plan of running the in-person study into a fully-remote one due to the COVID-19 outbreak. Here, we share some of the challenges and issues we encountered and how we alleviated them. First, it was infeasible to effectively demonstrate multimodal interactions (e.g., the push-to-talk recording) during the remote tutorial. We therefore prepared a video clip with subtitles and played it during the tutorial to introduce interaction methods to our participants. Second, we did not have control over participants' environment. Before running the pilot sessions, we sent a checklist to our participants (e.g., turning off the smartphone notification, connecting the laptop to a power cable) and asked them to follow the instructions. However, some participants did not actually comply with the provisions, even if they confirmed that they did. Furthermore, participants were occasionally distracted with pets or family members. To alleviate such issues in the main study sessions, we thoroughly checked whether participants turned on the most strict do-not-disturb mode prior to the screen sharing. Also, when participants were interrupted, we paused the session and asked them to handle the situation (e.g., closing the room door).

On the contrary, we were pleasantly surprised by unexpected advantages that our original in-person protocol would not have. First, because we were less constrained by time and location, we could reach a broad audience with diverse backgrounds and occupations. Second, the screen sharing app enabled us to record the smartphone screen in high-resolution, supporting better observations. Third, the research team members from remote locations could attend and observe the study session without much interference (by turning off the webcam and muting the microphone). We demonstrated that a remote study can be a viable option for deploying and testing multimodal interactions in a mobile app and hope that this study could inform other researchers wanting to design and run similar types of remote studies.

7 CONCLUSION

We presented Data@Hand, a novel mobile app that combines two complementary input modalities, speech and touch, to support exploring personal health data on smartphones. Data@Hand supports three types of interactions—touch-only, speech-only, and touch+speech—to enable flexible time manipulation, temporal comparisons, and data-driven queries. To examine how multimodal interaction helps people explore their own data, we conducted an exploratory think-aloud study with 13 long-term Fitbit users. Participants successfully adopted multimodal interaction and used both speech and touch interactions while finding personal insights. We also learned when and why people choose one interaction modality over others. We highlighted several areas for future research, including incorporating personally meaningful events and contexts, improving the recognition and interpretation of speech commands, and refining visualizations for further enhancing data exploration. We also showed that a remote study can be a viable option for deploying and testing a mobile app with multimodal interaction. In summary, our work contributes the first mobile app that leverages the synergy of speech and touch input modalities for personal data exploration, and the study conducted with participants' own long-term data on their devices. We hope this work can inform and inspire researchers in the visualization and broad HCI communities to leverage multimodal interactions to foster fluid and flexible data exploration on smartphones.

ACKNOWLEDGMENTS

We thank our study participants for their time and efforts. We also thank Niklas Elmqvist, Snehesh Shrestha, and the HCIL members at the University of Maryland at College Park, who provided feedback on the early version of this paper. Jarrett Lee, Lily Huang, Rachael Zehrung, Yuhan Luo, Pramod Chundury, and Ignacio Jauregui also helped us improve the tutorial protocol and material.

This study was in part supported by the National Science Foundation award #1753452. Young-Ho Kim was in part supported by Basic Science Research Program through the National Research Foundation in Korea, funded by the Ministry of Education (NRF-2019R1A6A3A12031352).

REFERENCES

[1] Christopher Ahlberg and Ben Shneiderman. 2003. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *The Craft of Information Visualization*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 7–13. https://doi.org/10.1145/191666.191775

[2] Majedah Alrehiely, Parisa Eslambolchilar, and Rita Borgo. 2018. A Taxonomy for Visualisations of Personal Physical Activity Data on Self-Tracking Devices and Their Applications. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. BCS Learning and Development Ltd., Swindon, UK, 1–15. https://doi.org/10.14236/ewic/HCI2018.17

[3] Apple Inc. 2021. Health. Retrieved Jan 05, 2021 from http://www.apple.com/ios/health/

[4] Apple Inc. 2021. SIRI shortcuts boost health and fitness routines - Apple News Room. Retrieved Jan 05, 2021 from https://www.apple.com/newsroom/2019/03/siri-shortcuts-boost-health-and-fitness-routines/

[5] Apple Inc. 2021. Speech | Apple Developer. Retrieved Jan 05, 2021 from https://developer.apple.com/documentation/speech

[6] Bon Adriel Aseniero, Charles Perin, Wesley Willett, Anthony Tang, and Sheelagh Carpendale. 2020. Activity River: Visualizing Planned and Logged Personal Activities for Reflection. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces (AVI '20)*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3399715.3399921

[7] Jillian Aurisano, Abhinav Kumar, Alberto Gonzales, Khairi Reda, Jason Leigh, Barbara Di Eugenio, and Andrew Johnson. 2015. "Show me data": Observational Study of a Conversational Interface in Visual Data Exploration. In *IEEE VIS 2015 (Poster)*. IEEE, Washington, DC, USA, 2 pages.

[8] Jillian Aurisano, Abhinav Kumar, Alberto Gonzalez, Jason Leigh, Barbara Di Eugenio, and Andrew Johnson. 2016. Articulate2: Toward a Conversational Interface for Visual Data Exploration. In *IEEE VIS 2016 (Poster)*. IEEE, Washington, DC, USA, 3–4.

[9] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, and Petra Isenberg. 2018. Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 630–640. https://doi.org/10.1109/TVCG.2018.2865142

[10] Kerstin Blumenstein, Christina Niederer, Markus Wagner, Grischa Schmiedl, Alexander Rind, and Wolfgang Aigner. 2016. Evaluating Information Visualization on Mobile Devices. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16*, Vol. 24-October. ACM Press, New York, New York, USA, 125–132. https://doi.org/10.1145/2993901.2993906

[11] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2018. Visualizing Ranges Over Time on Mobile Phones: a Task-Based Crowdsourced Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 619–629. https://doi.org/10.1109/TVCG.2018.2865234

[12] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2019. A Comparative Evaluation of Animation and Small Multiples for Trend Visualization on Mobile Phones. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 364–374. https://doi.org/10.1109/TVCG.2019.2934397

[13] Yang Chen. 2017. Visualizing Large Time-series Data on Very Small Screens. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*. The Eurographics Association, Goslar, Germany, 37–41. https://doi.org/10.2312/eurovisshort.20171130

[14] Eun Kyoung Choe, Raimund Dachselt, Petra Isenberg, and Bongshin Lee. 2019. Mobile Data Visualization (Dagstuhl Seminar 19292). *Dagstuhl Reports* 9 (2019). Issue 7.

[15] Eun Kyoung Choe, Bongshin Lee, and Seung-won Hwang. 2018. Personal Data Exploration with Speech on Mobile Devices. In *MultimodalVis '18 Workshop at AVI 2018*. ACM, New York, NY, USA, 4 pages. https://www.microsoft.com/en-us/research/publication/personal-data-exploration-with-speech-on-mobile-devices/

[16] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. 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* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 121–132. https://doi.org/10.1145/2750858.2804266

[17] Eun Kyoung Choe, Bongshin Lee, and M.c. Schraefel. 2015. Characterizing Visualization Insights from Quantified Selfers' Personal Data Presentations. *IEEE Computer Graphics and Applications* 35, 4 (July 2015), 28–37. https://doi.org/10.1109/MCG.2015.51

[18] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. 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* (Barcelona, Spain) *(PervasiveHealth '17)*. ACM, New York, NY, USA, 173–182. https://doi.org/10.1145/3154862.3154881

[19] Philip R Cohen, Mary Dalrymple, Douglas B Moran, Fernando C N Pereira, and Joseph W Sullivan. 1989. Synergistic Use of Direct Manipulation and Natural Language. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '89)*. ACM, New York, NY, USA, 227–233. https://doi.org/10.1145/67450.67494

[20] Joëlle Coutaz, Laurence Nigay, Daniel Salber, Ann Blandford, Jon May, and Richard M. Young. 1995. Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The Care Properties. In *Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction (INTERACT '95)*. Springer US, Boston, MA, 115–120. https://doi.org/10.1007/978-1-5041-2896-4_19

[21] Kenneth Cox, Rebecca E. Grinter, Stacie L. Hibino, Lalita Jategaonkar Jagadeesan, and David Mantilla. 2001. A Multi-Modal Natural Language Interface to an Information Visualization Environment. *International Journal of Speech Technology* 4, 3-4 (2001), 297–314. https://doi.org/10.1023/A:1011368926479

[22] Pooja M Desai, Matthew E Levine, David J Albers, and Lena Mamykina. 2018. Pictures Worth a Thousand Words: Reflections on Visualizing Personal Blood Glucose Forecasts for Individuals with Type 2 Diabetes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174112

[23] Kedar Dhamdhere, Kevin S. McCurley, Ralfi Nahmias, Mukund Sundararajan, and Qiqi Yan. 2017. Analyza: Exploring Data with Conversation. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM Press, New York, New York, USA, 493–504. https://doi.org/10.1145/3025171.3025227

[24] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction* 31, 4 (April 2015), 307–335. https://doi.org/10.1080/10447318.2014.986642

[25] Philipp Eichmann, Darren Edge, Nathan Evans, Bongshin Lee, Matthew Brehmer, and Christopher White. 2020. Orchard: Exploring Multivariate Heterogeneous Networks on Mobile Phones. *Computer Graphics Forum* 39, 3 (June 2020), 115–126. https://doi.org/10.1111/cgf.13967

[26] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 667–676. https://doi.org/10.1145/2598510.2598558

[27] Facebook Inc. 2021. React Native · A framework for building native apps using React. Retrieved Jan 05, 2021 from https://reactnative.dev/

[28] Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. 2018. 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 (2018), 1–21. https://doi.org/10.1145/3264917

[29] Fitbit, Inc. 2021. Aria Scale. Retrieved Jan 05, 2021 from https://www.fitbit.com/aria

[30] Fitbit, Inc. 2021. Fitbit. Retrieved Jan 05, 2021 from https://www.fitbit.com

[31] Fitbit, Inc. 2021. Fitbit Web API. Retrieved Jan 05, 2021 from https://dev.fitbit.com/build/reference/web-api/

[32] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM Press, New York, New York, USA, 489–500. https://doi.org/10.1145/2807442.2807478

[33] Garmin Ltd. 2021. Garmin. Retrieved Jan 05, 2021 from https://garmin.com

[34] Marti Hearst and Melanie Tory. 2019. Would You Like A Chart with That? Incorporating Visualizations into Conversational Interfaces. In *2019 IEEE Visualization Conference, VIS 2019*. IEEE, Washington, DC, USA, 36–40. https://doi.org/10.1109/VISUAL.2019.8933766

[35] Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. 2017. Applying Pragmatics Principles for Interaction with Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 309–318. https://doi.org/10.1109/TVCG.2017.2744684

[36] Huami Inc. 2021. Mi Fit. Retrieved Jan 05, 2021 from https://apps.apple.com/us/app/mi-fit/id938688461

[37] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2014. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (2014), 420–433. https://doi.org/10.1109/TVCG.2014.2359887

[38] Dandan Huang, Melanie Tory, and Lyn Bartram. 2016. A Field Study of On-Calendar Visualizations. In *Proceedings of Graphics Interface 2016* (Victoria, British Columbia, Canada) *(GI 2016)*. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, Waterloo, Canada, 13–20. https://doi.org/10.20380/GI2016.03

[39] Jan-Frederik Kassel and Michael Rohs. 2018. Valletto: A Multimodal Interface for Ubiquitous Visual Analytics. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188445

[40] Jan-Frederik Kassel and Michael Rohs. 2019. Talk to Me Intelligibly: Investigating An Answer Space to Match the User's Language in Visual Analysis. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 1517–1529. https://doi.org/10.1145/3322276.3322282

[41] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A. Kientz. 2012. Lullaby: A Capture & Access System for Understanding the Sleep Environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(Ubi-Comp '12)*. ACM, New York, NY, USA, 226–234. https://doi.org/10.1145/2370216.2370253

[42] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5092–5103. https://doi.org/10.1145/2858036.2858558

[43] Spencer Kelly. 2019. Compromise. Retrieved Jan 05, 2021 from http://compromise.cool/

[44] Yoojung Kim, Bongshin Lee, and Eun Kyoung Choe. 2019. Investigating data accessibility of personal health apps. *Journal of the American Medical Informatics Association* 26, 5 (May 2019), 412–419. https://doi.org/10.1093/jamia/ocz003

[45] Yea-Seul Kim, Mira Dontcheva, Eytan Adar, and Jessica Hullman. 2019. Vocal Shortcuts for Creative Experts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM Press, New York, New York, USA, 1–14. https://doi.org/10.1145/3290605.3300562

[46] Konstantin Klamka, Tom Horak, and Raimund Dachselt. 2020. Watch+Strap: Extending Smartwatches with Interactive StrapDisplays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376199

[47] Abhinav Kumar, Barbara Di Eugenio, Jillian Aurisano, Andrew Johnson, Abeer Alsaiari, Nigel Flowers, Alberto Gonzalez, and Jason Leigh. 2017. Multimodal Coreference Resolution for Exploratory Data Visualization Dialogue: Context-Based Annotation and Gesture Identification. In *SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue*. ISCA, Saarbrücken, Germany, 41–51. https://doi.org/10.21437/SemDial.2017-5

[48] Bongshin Lee, Matthew Brehmer, Petra Isenberg, Eun Kyoung Choe, Ricardo Langner, and Raimund Dachselt. 2018. Data Visualization on Mobile Devices. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. ACM, New York, NY, USA, Article W07, 8 pages. https://doi.org/10.1145/3170427.3170631

[49] Bongshin Lee, Eun Kyoung Choe, Petra Isenberg, Kim Marriott, John Stasko, and Theresa-Marie Rhyne. 2020. Reaching Broader Audiences With Data Visualization. *IEEE Computer Graphics and Applications* 40, 2 (March 2020), 82–90. https://doi.org/10.1109/MCG.2020.2968244

[50] Bongshin Lee, Arjun Srinivasan, John Stasko, Melanie Tory, and Vidya Setlur. 2018. Multimodal Interaction for Data Visualization. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. Association for Computing Machinery, New York, NY, USA, Article 11, 3 pages. https://doi.org/10.1145/3206505.3206602

[51] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. ACM, New York, NY, USA, 405–414. https://doi.org/10.1145/2030112.2030166

[52] Jean-claude Martin. 1998. TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. In *Intelligence and Multimodality in Multimedia Interfaces*. AAAI Press, Palo Alto, CA, USA, 1–25.

[53] Microsoft. 2021. Cognitive Speech Services | Microsoft Azure. Retrieved Jan 05, 2021 from https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/

[54] Microsoft. 2021. TypeScript. Retrieved Jan 05, 2021 from https://www.typescriptlang.org

[55] Laura Pina, Sang-Wha Sien, Clarissa Song, Teresa M Ward, James Fogarty, Sean A Munson, and Julie A Kientz. 2020. DreamCatcher: Exploring How Parents and School-Age Children Can Track and Review Sleep Information Together. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–25. https://doi.org/10.1145/3392882

[56] Irene Ros. 2021. MobileVis. Retrieved Jan 05, 2021 from http://mobilev.is/

[57] Sebastian Sadowski. 2021. Mobile InfoVis. Retrieved Jan 05, 2021 from https://mobileinfovis.com

[58] Ayshwarya Saktheeswaran, Arjun Srinivasan, and John Stasko. 2020. Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis. *IEEE Transactions on Visualization and Computer Graphics* 0, 0 (2020), 2168–2179. (In press).

[59] Samsung. 2021. S Health. Retrieved Jan 05, 2021 from http://shealth.samsung.com

[60] Hanna Schneider, Julia Wayrauther, Mariam Hassib, and Andreas Butz. 2019. Communicating Uncertainty in Fertility Prognosis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300391

[61] Michail Schwab, Sicheng Hao, Olga Vitek, James Tompkin, Jeff Huang, and Michelle A Borkin. 2019. Evaluating Pan and Zoom Timelines and Sliders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300786

[62] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 365–377. https://doi.org/10.1145/2984511.2984588

[63] Vidya Setlur, Melanie Tory, and Alex Djalali. 2019. Inferencing Underspecified Natural Language Utterances in Visual Analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 40–51. https://doi.org/10.1145/3301275.3302270

[64] Arjun Srinivasan, Bongshin Lee, Nathalie Henry Riche, Steven M. Drucker, and Ken Hinckley. 2020. InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376782

[65] Arjun Srinivasan, Bongshin Lee, and John T. Stasko. 2020. Interweaving Multimodal Interaction with Flexible Unit Visualizations for Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 0, 0 (2020), 15 pages. https://doi.org/10.1109/TVCG.2020.2978050 (Early access).

[66] Arjun Srinivasan and John Stasko. 2017. Natural Language Interfaces for Data Analysis with Visualization: Considering What Has and Could Be Asked. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers* (Barcelona, Spain) *(EuroVis '17)*. The Eurographics Association, Goslar, Germany, 55–59. https://doi.org/10.2312/eurovisshort.20171133

[67] Arjun Srinivasan and John Stasko. 2018. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (Jan. 2018), 511–521. https://doi.org/10.1109/TVCG.2017.2745219

[68] Poorna Talkad Sukumar, Gonzalo J. Martinez, Ted Grover, Gloria Mark, Sidney K. D'Mello, Nitesh V. Chawla, Stephen M. Mattingly, and Aaron D. Striegel. 2020. Characterizing Exploratory Behaviors on a Personal Visualization Interface Using Interaction Logs. In *EuroVis 2020 - Short Papers*, Andreas Kerren, Christoph Garth, and G. Elisabeta Marai (Eds.). The Eurographics Association, Goslar, Germany, 1–5. https://doi.org/10.2312/evs.20201052

[69] Yiwen Sun, Jason Leigh, Andrew Johnson, and Sangyoon Lee. 2010. Articulate: A Semi-Automated Model for Translating Natural Language Queries into Meaningful Visualizations. In *International Symposium on Smart Graphics*. Springer-Verlag, Berlin, Heidelberg, 184–195.

[70] Wanasit Tanakitrungruang. 2014. Chrono. Retrieved Jan 05, 2021 from https://github.com/wanasit/chrono

[71] TeamViewer. 2021. TeamViewer QuickSupport. Retrieved Jan 05, 2021 from https://www.teamviewer.com/en-us/info/quicksupport/

[72] Alice Thudt, Dominikus Baur, Samuel Huron, and Sheelagh Carpendale. 2016. Visual Mementos: Reflecting Memories with Personal Data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 369–378. https://doi.org/10.1109/TVCG.2015.2467831

[73] Alice Thudt, Bongshin Lee, Eun Kyoung Choe, and Sheelagh Carpendale. 2017. Expanding Research Methods for a Realistic Understanding of Personal Visualization. *IEEE Computer Graphics and Applications* 37, 2 (2017), 12–18. https://doi.org/10.1109/MCG.2017.23

[74] Viktoriya Trifonova. 2018. How Device Usage Changed in 2018 and What it Means for 2019. https://blog.globalwebindex.com/trends/device-usage-2019.

[75] Matthew Turk. 2014. Multimodal Interaction: A Review. *Pattern Recognition Letters* 36, 15 (Jan. 2014), 189–195. https://doi.org/10.1016/j.patrec.2013.07.003

[76] Andries van Dam. 2001. Post-Wimp User Interfaces: the Human Connection. In *Frontiers of Human-Centered Computing, Online Communities and Virtual Environments*. Springer London, London, 163–178. https://doi.org/10.1007/978-1-4471-0259-5_11

[77] Minh Tue Vo and Alex Waibel. 1993. Multimodal Human-Computer Interaction. In *Proceedings of ISSD 1993 (ISSD '93)*. Waseda University, Shinjuku City, Tokyo,

Japan, 1–8.
[78] Benjamin Watson and Vidya Setlur. 2015. Emerging Research in Mobile Visualization. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 883–887. https://doi.org/10.1145/2786567.2786571
[79] Bowen Yu and Claudio T. Silva. 2020. FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (Jan. 2020), 1–11. https://doi.org/10.1109/TVCG.2019.2934668
[80] Zoom Video Communications, Inc. 2021. Zoom. Retrieved Jan 05, 2021 from https://zoom.us/