FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input

Yuhan Luo
yuhanluo@umd.edu
University of Maryland
College Park, MD, USA

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

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

Naeemul Hassan
nhassan@umd.edu
University of Maryland
College Park, MD, USA

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

ABSTRACT

The factors influencing people's food decisions, such as one's mood and eating environment, are important information to foster self-reflection and to develop personalized healthy diet. But, it is difficult to consistently collect them due to the heavy data capture burden. In this work, we examine how speech input supports capturing everyday food practice through a week-long data collection study (N = 11). We deployed FoodScrap, a speech-based food journaling app that allows people to capture food components, preparation methods, and food decisions. Using speech input, participants detailed their meal ingredients and elaborated their food decisions by describing the eating moments, explaining their eating strategy, and assessing their food practice. Participants recognized that speech input facilitated self-reflection, but expressed concerns around re-recording, mental load, social constraints, and privacy. We discuss how speech input can support low-burden and reflective food journaling and opportunities for effectively processing and presenting large amounts of speech data.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Sound-based input / output**; **Field studies**.

KEYWORDS

Food tracking, self-tracking, personal informatics, speech input, speech interface design

ACM Reference Format:
Yuhan Luo, Young-Ho Kim, Bongshin Lee, Naeemul Hassan, and Eun Kyoung Choe. 2021. FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28-July 2, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3461778.3462074

1 INTRODUCTION

Food journaling supports a variety of health goals such as weight loss and balanced diet [24]. In the digital era, we see numerous technologies that support food journaling, including photo [42], barcode scanning [5], accelerated search [35], and smart sensors [46]. While these technologies predominantly focus on capturing calories and nutrients, researchers have highlighted the importance of capturing relevant factors that play parts in people's food practice (e.g., time of eating, mood, eating environments), which are essential for individuals to perform self-reflection [23, 73] and for health professionals to make personalized diet recommendations [18, 44]. Because food practice is highly individualized, it is difficult to capture "unified" key factors that influence everyone's food practice with automated approaches [44, 65]. Free-form text input allows individuals to describe their food practice in a flexible manner, but can impose heavy data capture burden [9].

In recent years, speech interaction has been growing in popularity with the introduction of voice assistants, such as Google Assistant [32], Amazon Alexa [4], and Apple Siri [6]. Because people speak faster than they type [56], researchers have begun to build data collection tools leveraging speech input to increase the effectiveness [25, 45, 59]. In addition, people tend to be expressive when they speak [13]: for example, in previous survey studies, participants who used speech input provided more elaborated answers than those who used text [45, 55]. Hence, we see the potential of speech input for collecting rich details while lowering the data capture burden.

However, despite speech input's potential for fast and expressive data capture, we have little knowledge on how it can be useful in capturing unstructured self-tracking data such as food practice. In this light, we examine how speech input can support capturing everyday food practice by deploying a food journaling app called FoodScrap, which captures food components, preparation methods, and food decisions in free-form audio recordings. In particular, we designed four guided prompts asking why people decide *when to eat*, *what to eat*, *how much to eat*, and *when they make the decision*, which are key questions in examining the multifaceted aspects in food decision-making [8, 65]. Although understanding how people make food decisions has long been an interest in food science research, a majority of prior work employed questionnaires and interviews to retrospectively identify factors that influence food

DIS '21, June 28-July 2, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8476-6/21/06…$15.00
https://doi.org/10.1145/3461778.3462074

decisions [7, 10, 20, 60] rather than examine how people make their food decisions *in-situ.*

With FoodScrap, we aim to understand the experience of capturing everyday food practice using speech input, focusing on *data richness* (i.e., the amount of data and the level of details) and *data capture burden* (i.e., how easy or difficult to capture data). Such understanding could help us envision how we can incorporate speech input in self-tracking tools. We conducted a one-week data collection study deploying FoodScrap to 11 participants who were interested in understanding their food decisions but were not practicing food journaling at the time of study. After the data collection, we measured participants' perceived data capture burden using a set of subscales from User Burden Scale (UBS) [66], followed by debriefing interviews.

In the study, participants produced rich data around their food practices. Not only did they detail the ingredients in their meals and steps of preparation procedures, but they also elaborated their food decisions by describing the eating moments, explaining their eating strategy, and assessing their food practice. Participants reported a low perceived user burden while expressing concerns around re-recording, mental load, social constraints, and privacy. Although we deployed FoodScrap mainly as a data collection tool, participants recognized its speech input as a way to facilitate self-reflection. During the debriefing interviews, participants discussed the insights they learned from capturing everyday food decisions and how they reflected on their decisions during the moment of data capture. Drawing from the findings, we distill how FoodScrap enabled situated reflection with speech input and the guided prompts. To fully leverage the rich data collected by speech input, we discuss potential solutions to process the data and to deliver meaningful visual and auditory feedback. Furthermore, reflecting on the benefits and drawbacks of speech input, we discuss opportunities for designing multimodal self-tracking technologies to support food journaling in different scenarios.

The contributions of this work are threefold. First, we provide an empirical understanding on how speech input supports people capturing unstructured food practice data on mobile devices, including what information people capture via free-form speech and how much data capture burden they perceive. Second, based on the rich information our participants collected, we offer implications for effectively processing and presenting large amounts of speech data. Third, we inform the design of multimodal self-tracking systems to support low-burden and reflective food journaling.

2 RELATED WORK

In this section, we cover related work in the areas of (1) multimodal food journaling approaches focusing on capturing various aspects in food practice and (2) speech-based data collection.

2.1 Multimodal Food Journaling

Food journaling has become a prevalent approach for individuals to monitor their diet [12], but is also known to be burdensome due to the complexity of meal composition and variation in preparation methods [24]. In the Human-Computer Interaction (HCI) community, much effort has been make to lower the burden of food journaling through different input modalities, including photo-based food journals [42], barcode scanning [5], accelerated search [35], and smart sensors [50]. Among these input modalities, photo-based food journal is most popular for its convenience [42] and ease of sharing [17]. Because food photos may cover information such as location and social elements, they further reduce the burden of capturing additional eating contexts [23]. However, food photos cannot always capture necessary details such as portion size, condiments, and individual ingredients, especially when the meal preparation methods are complicated [9, 18]. In addressing such challenges, Chung and colleagues designed Foodprint by enabling people to add contextual information (e.g., mood, symptoms, free-form text descriptions) in addition to food photos to aid reflection and data sharing [18]. In addition, researchers have built systems to collect food data with automated methods [46, 50]. For example, Mirtchouk and colleagues developed a body-worn wearable device that can recognize food types and quantities with motion and audio sensors [46]. Although their approach reduced input burden and improved data accuracy, it may undermine the benefits of in-the-moment awareness created by food journaling.

Leveraging smartphones' voice assistants, both commercial apps (e.g., Talk-to-Track [31]) and research prototypes [41, 62, 63] have incorporated speech to support food journaling. For example, Korpusik and colleagues developed Coco Nutrition [40], a conversational calorie counter that allows people to describe their food intake with speech input and automatically calculates their calorie consumption [41]. Their work focused on improving the accuracy of food recognition in natural languages and matching users' input with food entries in the USDA database [41]. In addition, in Silva and colleagues' work in progress, they developed a multimodal food journal across multiple platforms including smartphones, web browsers, Google Home, Amazon Alexa, and Apple Watch [62]. Although their work overlaps with ours in examining how people capture their food data using speech input, their work focused on how people choose among different devices for data capture, while our work focuses on the level of detail that people capture.

While prior work primarily focused on capturing food components with the aim of providing more accurate nutrients and calorie information, we see a growing interest in Human-Food Interaction (HFI), which focuses on enriching food practice ranging from how people cook, how they interact with food, and how food influence their daily life [3, 37]. In particular, the practice of food journaling has been expanded to capture broader eating contexts (e.g., mood, eating environment) beyond what people eat [9, 68, 73]. For example, to promote self-reflection on one's food practice, Zhang and colleagues developed Eat4Thought, which captures a variety of eating contexts such as mood, emotion, and eating environment using video recording [73]. To further leverage the large amount of contextual information, Terzimehić and colleagues collected a rich set of food choice moments, which informed design opportunities for just-in-time adaptive interventions (JITAI) to encourage healthy eating [68]. For the ease of data analysis, these work often predefined certain structures in terms of what data to capture and which format to use (e.g., selecting a mood from existing options), but little work has looked into what people capture about their food decisions in unstructured forms, how rich the information is, and how much data capture burden it imposes to people.

2.2 Speech-Based Data Collection

With the rise of speech recognition and natural language processing (NLP) technologies, speech has become a prevalent modality to interact with digital devices [19]. Because people speak faster than they type [56], researchers have explored the opportunities for speech-based data collection [21, 25, 55, 59]. For example, Revilla and colleagues compared speech with text input in responding to survey questions, and found that participants who used speech input spent less time and provided more elaborated answers than those who used text [55]. In Patnaik and colleagues' survey study that collected individuals' health data related to tuberculosis, they found that speech input provided the more accurate information compared with SMS messages and electronic forms [52].

Another field that has explored speech-based data capture is clinical data entry, where doctors have to enter patients' data such as medical reports or prescriptions during or after meeting with patients [1, 26, 51, 58, 70, 72]. Researchers have developed speech recognition systems for clinical settings by incorporating medical terminology into the system vocabulary [1, 51, 58, 72]. In particular, Wenzel and colleagues showed that compared with handwriting and keyboard typing, doctors rated speech dictation with a higher level of satisfaction [72].

In addition, researchers suggested that speech input can effectively collect personal health data in self-tracking contexts [43, 44]. In Luo and colleagues' co-design study, dietitians brought up the idea of capturing one's feelings about food using speech input, which could encourage eating disorder patients to record frank thoughts without feeling shamed, because they do not need to review the captured information [44]. In another study where researchers studied how a smart speaker can complement a mobile app in exercise training and tracking, they found that the hands-free speech interaction made it easier than mobile app for people to capture their workout repetitions, especially when they were doing hands-intensive exercise (i.e., push-up) [43]. As a result, people were able to focus on their workout performance rather than worrying about having their smartphones close by [43]. Although their findings suggested speech input's potential for capturing short and structured data (i.e., number of workout repetitions) [43], there is a lack of empirical understanding on how speech input can benefit individuals in capturing unstructured and context-dependent self-tracking data. To bridge the research gap, we set out to incorporate speech input in context of food journaling, where capturing detailed food information and various eating contexts are important. With speech input's fast and expressive nature, we aim to lower the data capture burden while capturing rich information.

3 FOODSCRAP

We deployed a mobile food journaling app called FoodScrap with OmniTrack for Research[1], a web-based research tool that enables the creation and deployment of a flexible mobile self-tracking app [39]. The goal of this study is to examine the data richness and data capture burden of speech in capturing everyday food practice. Therefore, we designed FoodScrap as a data collection instrument without providing detailed feedback.

3.1 Journal Design

FoodScrap consists of three food journals: Main Meal Journal, Snack Journal, and Skip Journal. Figure 1 illustrates the interface of Main Meal Journal. All questions included in the journals were required. The logging time and session timestamps were automatically captured. The Main Meal Journal captures the following information for each meal:

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

Q2. Eating duration (start and end time)

Q3. A photo of the meal

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

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

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

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

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

Specifically, we asked people to take a food photo in Q3 so that they can remember to log their meals later.To ensure sure that people capture their meals close to the time they eat, Q2 only takes a time range that falls within the current day. Drawing from prior literature in food science [8, 65], we broke down food decisions into four aspects regarding when to eat, what to eat, when to make the decision, and how much to eat. As such, we designed questions Q5 to Q8 (Figure 1b), which serve as guided prompts to elicit the key aspects in food decision-making. The questions take free-form audio recordings as responses, providing the flexibility for people to express additional thoughts.

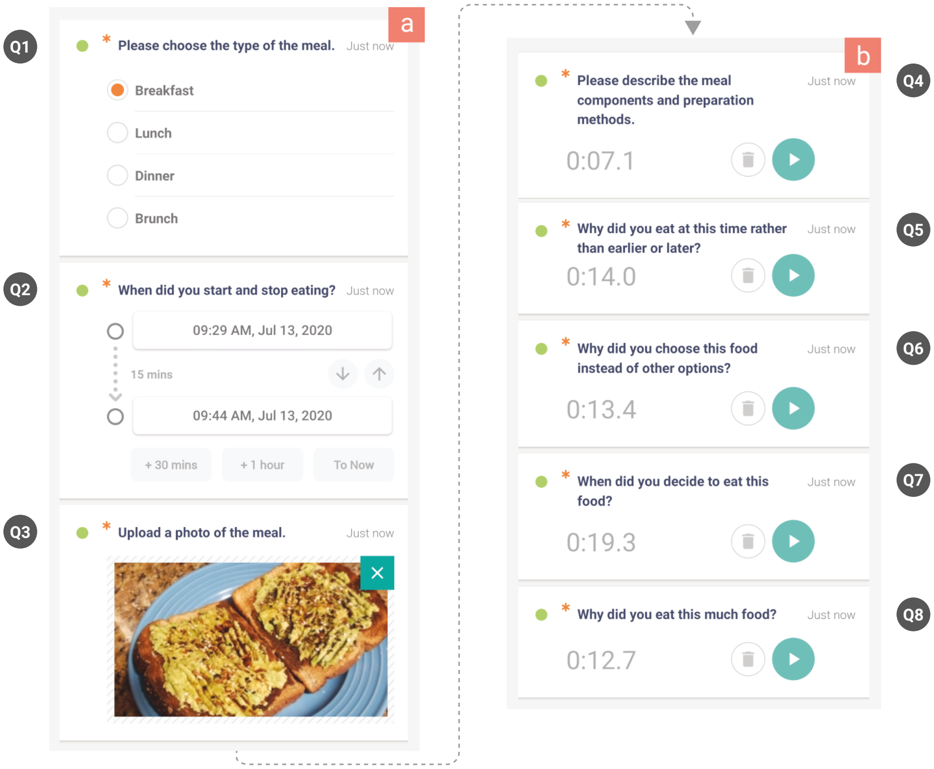
Snack Journal asks the same information as Main Meal Journal, except for Q1 (meal type). In addition, we designed Skip Journal to capture the main meals that people skip (excluding snacks) with three questions: the type of the meal that was skipped (SK1); *"When did you decide to skip the meal?"* (SK2);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

FoodScrap follows the design of commonly used voice recording interfaces (e.g., Samsung Voice Recorder [57]), which allows people to pause and resume the recording process. When recording is complete, people can play back their recording or delete the recording to start over. To exclude the effect of speech recognition errors that might influence user experience [1], we did not provide dictation or transcription support for speech input.

3.2 Daily Reminders

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

[1] https://omnitrack.github.io/research



Q1 * Please choose the type of the meal.

● Breakfast
○ Lunch
○ Dinner
○ Brunch

Q2 * When did you start and stop eating?

09:29 AM, Jul 13, 2020
15 mins
09:44 AM, Jul 13, 2020

Q3 * Upload a photo of the meal.

Q4 * Please describe the meal components and preparation methods. 0:07.1

Q5 * Why did you eat at this time rather than earlier or later? 0:14.0

Q6 * Why did you choose this food instead of other options? 0:13.4

Q7 * When did you decide to eat this food? 0:19.3

Q8 * Why did you eat this much food? 0:12.7

Figure 1: The data capture screen of Main Meal Journal in FoodScrap: (a) questions on meal type, eating duration, and photo of the meal; (b) questions on meal components, preparation methods, and food decisions.

the end-of-day reminder. Journal entries were considered valid as long as they were submitted within the same day.

4 METHODS

We deployed FoodScrap for seven consecutive days and conducted a post-study survey and debriefing interviews. Due to the COVID-19 outbreak, we interacted with participants remotely via a Zoom video call [74] (in June-July 2020). The study was approved by the university’s Institutional Review Board. Unlike traditional self-tracking studies that focused on examining how tracking tools influenced participants’ tracking adherence [14] and behaviors [22], our work instead aimed at analyzing and understanding the nature of the captured information. Therefore, we structured our compensation to minimize missing journal entries without influencing the amount of data captured, which we describe in subsection 4.2.

4.1 Participants

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

We refined the study protocol and the FoodScrap design after working with the first participant, and excluded her data for later analysis. We excluded the data of two participants due to the data loss caused by technical issues. Therefore, we analyzed the data of the remaining 11 participants (P1–11; nine females and two males). Our participants lived in different regions in the US and their eating habits were influenced by diverse food cultures (See Table 1). Their age ranged from 18 to 60 (M = 30, SD = 11.40). Eight participants reported prior experience using speech input on their mobile phones. Although our participants were generally healthy individuals, they had specific eating goals such as eating healthier, losing weight, and reducing sweets intake. In particular, five participants reported struggling with food from time to time: P4 and P9 saw themselves as overweight, P8 and P9 thought they were sometimes emotional eaters, P10 was obsessed with sweets, and P6 tended to over exercise and had visited nutritionists regularly before the study. At the time of study, none of the participants were practicing food journaling.

4.2 Study Procedure

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

(1) Tutorial. We first had a one-on-one remote tutorial with each participant via a Zoom [74] video call (30 to 45 minutes). Participants were instructed to share their phone screen with us using TeamViewer QuickSupport [67], so that we could help them install FoodScrap in real-time. Before the screen sharing, we asked participants to remove any sensitive information from their home screen and to turn off all the notifications. We also shared our computer screen via Zoom, which allowed participants to see how their phone screen was displayed to us. During the tutorial, we introduced the study procedure and explained the information that participants needed to capture. We also played a video clip demonstrating how to log an entry in Main Meal Journal. In addition, we asked each participant to estimate their regular eating time for the three main meals. We then customized their reminder time according to individual's meal times right after the tutorial.

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

(3) Post-Study Survey. At the end of the data collection, we emailed each participant a post-study survey to measure their perceived data capture burden with FoodScrap. The survey included a set of subscales taken from the User Burden Scale (UBS) [66], which was developed to capture different types of user burden with computing systems and was later validated in many HCI studies (e.g., [36, 71]). Specifically, we employed four out of six constructs from UBS: difficulty to use, time and social burden, mental and emotional burden, and privacy burden. Refer to our supplementary material for the full list of questions we used.

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

4.3 Data Analysis

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

When analyzing the responses in journal entries, we separately analyzed the responses to *meal/snack components and preparation*

Table 1: Participants' demographic and eating goals.

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

methods (Q4) and responses to questions on food decisions (Q5 to Q8) in Main Meal Journal and Snack Journal. For the meal/snack components and preparation methods, two authors first independently coded a subset of 247 responses (57; 23%). After resolving discrepancies in coding through multiple sessions of discussion, the first author coded the remaining entries. We analyzed the *types of details* that participants provided, rather than the actual *content* of the information because we were interested in examining the ways participants captured their food using speech rather than the types of food they chose. We first conducted Thematic Analysis [11] on the responses to identify common types of details, which are listed in Table 2. We then revisited the responses by checking which types of details they contained.

For the remaining responses to the questions from Q5 to Q8, two authors first independently coded a subset of the 988 responses (168; 17%), and followed the same procedure as we analyzed meal/snack components and preparation methods. We categorized the responses into three groups: (1) unelaborated response, which answered the question without further explanation; (2) elaborated response, which answered the questions with explanation and examples; and (3) digression, which digressed from the original question (See Section 5.3 for details). This categorization follows prior work on analyzing open-ended survey responses [55, 64], which defined an elaborated response as "*additional descriptive information or explanation about a theme without introducing a new theme*" [55, 64]. We focused on examining *whether* and *how* participants elaborated their responses rather than identifying factors that influenced their food decisions.

We audio-recorded the debriefing interviews and transcribed them into text. We grouped the interview transcripts to answer the following questions: (1) What participants liked and disliked about using speech to capture food practice; (2) what participants' experience was like in capturing their everyday food practice and how they reflected on their food decisions; (3) how participants perceived their data capture with speech input.

5 RESULTS

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

5.1 Descriptive Statistics of Journal Entries

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

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

5.2 Describing Details of Meal Components and Preparation Methods

By analyzing how participants described their *meal/snack components and preparation methods* (Q4), we identified nine different types of detail: dish names, ingredient types, individual ingredient items, spices & sauces, food portion, food characteristics, preparation types, procedural methods, and additional contexts. Table 2 summarizes the types of detail with descriptions, example quotes, and the number of responses in participants' journal entries.

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

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

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

5.3 Elaboration and Digression in Capturing Food Decisions

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

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

5.3.2 Elaborated Response. Elaborated responses refer to valid answers with additional information that detailed the answers. While analyzing the elaborated responses, we found that participants elaborated their responses by describing the eating moment, explaining the eating strategy, and assessing their food practice. In the following, we summarize each elaboration type (See Table 4 for descriptions, example quotes, and the number of responses).

(1) Describing the eating moment. Participants expanded their responses by describing what had happened around the eating moment. The most common instances were **personal status** such as activities and feelings. In P4's statement in Table 4, for example, she recalled what she did before eating: "*took a long nap*," "*did a lot of work around the house*," and "*picked up my dog*," as well as how she felt: "I *was so tired*." Another common form was describing one's **food access**, especially when responding to "*Why did you choose this food instead of other options*?" Participants mentioned their food availability or constraints such as "*running out of groceries*" (P7) and "*leftover that needed to be eaten before it goes bad*" (P6). In addition, participants described how their food decisions were influenced by **social and environmental contexts**, such as people around them: "*because my mom [was] really really late, and I was actually really looking forward to this specialty from her*" (P9), and their eating environment: "*It's extraordinarily hot today in Colorado, and I have no desire to turn on the oven or stove*" (P8).

(2) Explaining the eating strategy. Participants made food decisions based on a set of eating strategies they had specifically planned for convenience, health, or special events. Some of these eating strategies were adopted from other people or media sources, and later became participants' health belief or eating habits. The most commonly mentioned strategy is **planning ahead**. In P10's statement in Table 4, for example, she described how she prepared

Table 2: Summary of participants' responses to meal/snack components and preparation methods (Q4) in the Main Meal Journal and Snack Journal by the type of details they provided (Note that a response can include more than one type of details).

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

Table 3: Responses to questions regarding food decisions (Q5 to Q8) in the Main Meal and Snack journals, categorized into unelaborated responses, elaborated responses, and digression.

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

a big meal for several days. In another of P10's responses, she also explained how COVID-19 affected her eating strategies for planning ahead: "*I'm in food deliveries because of COVID. I've had to modify my diet and eat stuff like sandwiches, because my produce only lists the first week of the food order, and I'm ordering every two to three weeks for limited contact.*" The second eating strategy involves participants' **health belief**. For example, in Table 4, P4 believed that eating between 12 to 7 p.m. can help with weight loss. In another example, P7 believed that his food was healthy because "*this is a mix of protein as well as fiber.*" In addition, participants also mentioned their **habits** including the time they usually ate, the food they regularly chose, and the amount they usually consumed.

(3) Self-assessment. Another type of elaboration is self-assessment—participants expanded responses by assessing their food decisions. One common form was to make **judgment** with positive or negative comments. For example, P10 commented on one of her snacks: "*I wanted something sweet after dinner. It's a bad habit that started [since] the last couple years.*" Similarly, P7 described his lunch as "junk food." On the other hand, participants compared their current food decisions with their regular routines regarding eating time, healthiness of the food, and food amount, etc. In Table 4, for example, P3 noted, "*I would say I eat a little bit more than I normally do,*" which we categorized as **comparison**.

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

5.4 Benefits of Speech-Based Food Journaling

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

Table 4: Summary of participants' responses to the four questions on food decisions (Q5 to Q8) in the Main Meal Journal and Snack Journal by the ways they elaborated their responses (Note that a response can be elaborated in several ways, and the elaboration types and subtypes are not mutually exclusive).

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

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

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

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

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

5.5 Data Capture Burden

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

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

5.5.2 Mental Load. Participants reported that journaling with speech input sometimes required extra attention and concentration, especially in two cases: when they ate mindlessly without clear answers to the questions or when they had a lot to say about their food decisions. Four participants mentioned that they felt difficulty in responding to the questions on food decisions because of mindless eating: "*Most of the time I found myself eating, and I couldn't really tell why, why I ate at this time, or why I chose this food. I felt it's hard to give an answer, it might be just an intuition, or like a habit, but I can't explain why.*" (P1). On the other hand, P5 and P7 often needed to think through and organize what they wanted to speak before recording their responses. To make sure that their responses were clear and concise, it usually took extra mental load: "*Because I don't want to record [an] audio for a minute or two, where I'm fumbling through my sentences. So I needed to gather my thoughts regarding what I need to say quickly. So initially, it was a little jam regarding what I wanted to say*" (P7).

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

5.5.4 Privacy Concerns. Three participants considered food practice to be private, and were concerned about their food decisions

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

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

6 DISCUSSION

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

6.1 Collecting Rich Details Through Fast and Expressive Data Capture

Our participants provided rich details in their food components and preparation methods, which could be laborious to capture via touch-based typing, as P3 explained while showing one of his journal entries: "*This is a 45.7 second recording that I did. Now imagine, if I need to type, that would be too much writing. I'll probably miss some data or try to cut corners with it.*" We note that many of the details—such as condiments and preparation procedure—are critical information for assessing meal healthiness [54, 61, 69] but are difficult to capture through retrospective surveys or even automated food recognition technologies (e.g., photo, barcode) [49]. In dietary assessment, for example, dietitians and nutritionists often employ dietary history method [69] and food frequency questionnaire (FFQ) [61], which ask for more than 90 items about one's food intake, covering details such as "*seasonings and flavorings*" and "*cooking methods*," but do not always produce accurate results due to the time lag [69]. Our study suggests that speech-enabled data collection can capture more details *in-situ* with lower data capture burden, which may improve the data accuracy [52]. However, to fully leverage the large amount of speech data, we need to consider how to efficiently process and present the data for healthcare providers' use. With the advances in natural language processing (NLP), we can extract food-related information (e.g., food group, portion size, ingredients) from the transcribed text [41] and support sorting & filtering the information based on providers' needs [44].

In addition, when answering questions on food decisions, participants often elaborated their responses, which resonates with prior work suggesting that people tend to be expressive when they are speaking [13]. These elaborated responses are usually ephemeral and momentary contexts—personal status, food access, and social and environmental contexts around the time of eating—that are valuable information for dietitians and food science researchers, but can be hard to capture through retrospective recall. For example, understanding how patient's living environment and social life shape their food decisions helps dietitians deliver more personalized care: dietitians may help patients restructure their eating environment instead of simply prescribing what to eat [33], or use food journal as an intervention to encourage mindful eating [2, 44]. While current practice of understanding food decisions often relies on verbal communication during clinical consultation [34], FoodScrap enabled participants to capture food decisions that are tied to every meal or snack, providing opportunities to capture rich details that might otherwise be overlooked.

Our findings demonstrated the potential of speech input to capture detailed food information and elaborated food decisions that are typically hard to capture through other approaches (e.g., typing, automated means, interviews). In this regard, speech input holds promises in other self-tracking contexts beyond food journaling (e.g., capturing perceived workout intensity and feelings in exercise tracking), where individuals and researchers can identify nuanced but important insights from one's daily activities [15, 28].

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

Self-tracking technologies support reflection in various ways [53]: providing real-time feedback (e.g., [38]) or augmenting manual data capture(e.g., [14, 44, 73]) can support reflection-in-action; and providing aggregated feedback of past behaviors (e.g., [16, 22, 23, 29]) can support reflection-on-action. In our study, FoodScrap mainly facilitated reflection-in-action at the time of data capture. Among participants' elaborated responses, we found several instances involving self-assessment with *judgements* or *comparison*, which were indicators of reflection-in-action [30]. Those reflective thoughts were likely resulted from the guided prompts in FoodScrap, which questioned participants to think about their food decisions in specific aspects such as when and how much to eat, and why they choose the food. Furthermore, we suspect that speech input might have nudged reflective thinking by supporting free-form expressions, as P-10 remarked that thinking aloud was like "*interacting with people*," which made her feel "*more accountable*" to explain her food decisions. This finding corroborates a previous study in which researchers found that video recording of eating experience with narration could promote self-reflection through contextualizing one's eating experiences [73]. Such free-form expression is important for people who struggle with food (e.g., eating disorder patients) to raise situated awareness and to build positive self-image [44].

As reflection-in-action happens during the moment of data capture, which is close to the time of eating, we see opportunities for encouraging mindful eating during these "critical reflection moments" [2, 44, 68]. For example, asking "*why do you want to eat now?*" may prompt people to think twice about their decisions and to be more mindful about whether their cravings are caused by hunger or boredom [44]. To understand how different modalities of data capture (e.g., speech recording, video recording) support reflection-in-action, future work remains to compare these modalities with traditional text input or other structured entry forms.

6.3 Enabling *Reflection-on-Action* Through Feedback of Past Behaviors

To fully support a reflective food journaling experience, it is important to enable reflection-on-action through delivering aggregated or summary feedback of past behaviors, so that individuals can stay engaged by reflecting on the patterns of their food practices [14, 22].

The focus of the FoodScrap study was on the data capture aspect, so we did not provide any feedback beyond the capability of replaying the audio. While the unstructured nature of speech input adds complexity to data processing and analyzing, we see opportunities for presenting the rich information in both visual and auditory forms.

We can summarize individuals' responses corresponding to each guided prompt as feedback. For example, the most commonly mentioned reasons that affect one's eating time, food choice, and amount of food consumption can be shown in text summarization using keyword extraction and term-frequency analysis. To further support individuals exploring their data, the keywords extracted can be visualized in a word cloud [48].

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

Four (36%) participants in our replayed their audio recordings and valued their recordings as resources for revisiting past eating episodes. This finding implies the potential of auditory feedback to support reminiscence and reflection, which can be important for those who track their mood, stress, and mindful thoughts [47]. We suspect that when people audio record short and structured data consisting of numbers or simple phrases, text summary or chart is a better form than auditory feedback for reviewing purposes. On the other hand, when people capture long and complicated information, retaining the original audio recording could be valuable [27], as it might contain unique contextual information that text transcription cannot provide, such as pitch, tone, and volume of the voice as well as background sounds. To enable more efficient audio searching, we can provide text summary (e.g., extracted key words) or visual feedback (e.g., photos) along with the original audio recording.

6.4 Supporting Data Capture in Varying Contexts Leveraging Multimodal Input

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

To support food tracking in varying contexts, we can leverage multimodal input combining speech, text, and photo across multiple devices (e.g., smartphones, smart speakers, wearable devices, wireless earphones) so that people can choose *when* to use *which* input modality. For example, in a privacy-sensitive situation (e.g., crowded place, office setting), people may choose text input on a smartphone; at home where the smartphone is not close by, people can use speech input on a smart speaker or wearable devices with the hands-free interaction [43]. In another case when people do not have enough time to capture all the information at once, they can take a food photo first, and add more details afterward using speech or text.

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

6.5 Study Limitations and Future Work

Although we aimed to recruit participants from diverse backgrounds (9 types of occupations) and food cultures, our small (N = 11) and female-dominated (82%) sample may limit the generalizability of our findings. Studying with a larger population would likely produce more diverse results (e.g., identifying new detail types or elaboration types).

As the first step to explore the feasibility of using speech input to capture unstructured self-tracking data in the context of everyday food practice, our work identified rich insights regarding how speech input facilitated data collection and how participants perceived the data capture burden. Going forward, an important next step is to develop a pipeline to effectively process and present the large amount of speech data while collaborating with healthcare professionals. Furthermore, we envision that the lessons learned from this study can be extended to broader self-tracking contexts beyond a predefined data type (e.g., short text) or domain (e.g., food journaling). As such, we plan to incorporate speech input into a customizable setting where people can track diverse data types and decide when to use speech. Our overarching goal is to realize effective multimodal self-tracking technologies to support people better track, engage, and reflect on their everyday health by leveraging the strengths of different input modalities, including the fast and expressive speech input.

7 CONCLUSION

We reported a week-long data collection study with FoodScrap, a speech-based food journaling app that we created to capture food components, preparation methods, and food decisions. Throughout the study, 11 participants collected rich data, including detailed information about their food intake and elaborated statements of their food decisions. We distilled the ways that participants used speech input to describe their food practice, and summarized speech input's benefits and drawbacks regarding data capture burden. We highlighted speech input's fast and expressive data capture in collecting flexible and nuanced details and its potential for fostering reflection-in-action. We also discuss opportunities for leveraging speech input to further support reflection-on-action, and designing multimodal input systems to facilitate data capture in varying contexts. In summary, our work contributes to an empirical understanding on how speech input supports capturing unstructured self-tracking data. We hope this work can inform and inspire other researchers working in the growing body of personal health informatics to design multimodal self-tracking tools that capture rich information, lower data capture burden, and promote self-reflection.

ACKNOWLEDGMENTS

We thank all the participants for their time and effort. We also thank the reviewers for their thoughtful feedback. This research was in part funded by National Science Foundation (#1753452) and the Research Improvement Grant from the College of Information Studies at the University of Maryland, College Park.

REFERENCES

[1] Mohd Khanapi Abd Ghani and Ika Novita Dewi. 2012. Comparing speech recognition and text writing in recording patient health records. In *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*. IEEE, 365–370. https://doi.org/10.1109/IECBES.2012.6498100

[2] Susan Albers. 2012. *Eating mindfully: How to end mindless eating and enjoy a balanced relationship with food*. New Harbinger Publications.

[3] Ferran Altarriba Bertran, Samvid Jhaveri, Rosa Lutz, Katherine Isbister, and Danielle Wilde. 2019. Making sense of human-food interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3290605.3300908

[4] Amazon.com, Inc. 2021. *Amazon Alexa*. Retrieved April 26, 2021 from https://www.amazon.com/b?node=21576558011

[5] Norman Azah Anir, Hairul Nizam, and Azmi Masliyana. 2008. The users perceptions and opportunities in Malaysia in introducing RFID system for Halal food tracking. *WSEAS Transactions on information science and applications* 5, 5 (2008), 843–852.

[6] Apple, Inc. 2021. *Siri*. Retrieved April 26, 2021 from https://www.apple.com/siri/

[7] Gastón Ares, Franco Mawad, Ana Giménez, and Alejandro Maiche. 2014. Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference* 31 (2014), 28–37. https://doi.org/10.1016/j.foodqual.2013.07.005

[8] Els Bilman, Ellen van Kleef, and Hans van Trijp. 2017. External cues challenging the internal appetite control system—overview and practical implications. *Critical reviews in food science and nutrition* 57, 13 (2017), 2825–2834. https://doi.org/10.1080/10408398.2015.1073140

[9] Johnna Blair, Yuhan Luo, Ning F Ma, Sooyeon Lee, and Eun Kyoung Choe. 2018. OneNote Meal: A photo-based diary study for reflective meal tracking. In *AMIA Annual Symposium Proceedings*, Vol. 2018. American Medical Informatics Association, 252. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371351/

[10] Marcela CC Bomfim, Sharon I Kirkpatrick, Lennart E Nacke, and James R Wallace. 2020. Food literacy while shopping: Motivating informed food purchasing behaviour with a situated gameful app. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376801

[11] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[12] Lora E Burke, Melanie Warziski, Terry Starrett, Jina Choo, Edvin Music, Susan Sereika, Susan Stark, and Mary Ann Sevick. 2005. Self-monitoring dietary intake: current and future practices. *Journal of Renal Nutrition* 15, 3 (2005), 281–290. https://doi.org/10.1016/j.jrn.2005.04.002

[13] Barbara L Chalfonte, Robert S Fish, and Robert E Kraut. 1991. Expressive richness: a comparison of speech and text as media for revision. In *Proceedings of the 1991 CHI Conference on Human Factors in Computing Systems*. 21–26. https://doi.org/10.1145/108844.108848

[14] 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*. 121–132. https://doi.org/10.1145/2750858.2804266

[15] 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 (2015), 28–37. https://doi.org/10.1109/MCG.2015.51

[16] 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*. 173–182. https://doi.org/10.1145/3154862.3154881

[17] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When personal tracking becomes social: Examining the use of Instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on human factors in computing systems*. 1674–1687. https://doi.org/10.1145/3025453.3025747

[18] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A Munson. 2019. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 1 (2019), 1–27. https://doi.org/10.1145/3314394

[19] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers* 31, 4 (2019), 349–371. https://doi.org/10.1093/iwc/iwz016

[20] Deborah A Cohen and Susan H Babey. 2012. Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity Reviews* 13, 9 (2012), 766–779. https://doi.org/10.1111/j.1467-789X.2012.01001.x

[21] Frederick G Conrad, Michael F Schober, Christopher Antoun, H Yanna Yan, Andrew L Hupp, Michael Johnston, Patrick Ehlen, Lucas Vickers, and Chan Zhang. 2017. Respondent mode choice in a smartphone survey. *Public Opinion Quarterly* 81, S1 (2017), 307–337. https://doi.org/10.1093/poq/nfw097

[22] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems*. 1797–1806. https://doi.org/10.1145/1357054.1357335

[23] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*. 3207–3216. https://doi.org/10.1145/2702123.2702154

[24] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. 2015. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*. 1159–1162. https://doi.org/10.1145/2702123.2702155

[25] Marika De Bruijne and Arnaud Wijnant. 2013. Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey. *Social Science Computer Review* 31, 4 (2013), 482–504. https://doi.org/10.1177/0894439313483976

[26] Yaron D Derman, Tamara Arenovich, and John Strauss. 2010. Speech recognition software and electronic psychiatric progress notes: physicians' ratings and preferences. *BMC medical informatics and decision making* 10, 1 (2010), 1–7. https://doi.org/10.1145/1497185.1497286

[27] Daniel PW Ellis and Keansub Lee. 2004. Minimal-impact audio-based personal archives. In *Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences*. 39–47. https://doi.org/10.1145/1026653.1026659

[28] 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*. 667–676. https://doi.org/10.1145/2598510.2598558

[29] 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

[30] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 216–223. https://doi.org/10.1145/1952222.1952269

[31] Genesant Technologies, Inc. 2021. Talk-to-Track. Retrieved April 26, 2021 from https://www.talktotrack.com/

[32] Google, Inc. 2021. *Google Assistant*. Retrieved April 26, 2021 from https://assistant.google.com/

[33] Mette Holst, Tina Beermann, Marie Nerup Mortensen, Lotte Boa Skadhauge, Marianne Køhler, Karen Lindorff-Larsen, and Henrik Højgaard Rasmussen. 2017. Optimizing protein and energy intake in hospitals by improving individualized meal serving, hosting and the eating environment. *Nutrition* 34 (2017), 14–20. https://doi.org/10.1016/j.nut.2016.05.011

[34] Amy Leung Hui, Gustaaf Sevenhuysen, Dexter Harvey, and Elizabeth Salamon. 2014. Barriers and coping strategies of women with gestational diabetes to follow dietary advice. *Women and Birth* 27, 4 (2014), 292–297. https://doi.org/10.1016/j.wombi.2014.07.001

[35] Jisu Jung, Lyndal Wellard-Cole, Colin Cai, Irena Koprinska, Kalina Yacef, Margaret Allman-Farinelli, and Judy Kay. 2020. Foundations for systematic evaluation and benchmarking of a mobile food logger in a large-scale nutrition study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–25. https://doi.org/10.1145/3397327

[36] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine

Zia. 2017. Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6850–6863. https://doi.org/10.1145/3025453.3025480

[37] Rohit Ashok Khot, Florian Mueller, et al. 2019. Human-food interaction. *Foundations and Trends® in Human-Computer Interaction* 12, 4 (2019), 238–415. http://dx.doi.org/10.1561/1100000074

[38] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 272–283. https://doi.org/10.1145/2858036.2858428

[39] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: A flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–28. https://doi.org/10.1145/3130930

[40] Mandy Korpusik and Jim Glass. 2021. Coco Nutrition. https://www.coco-nutrition.com/.

[41] Mandy Korpusik, Salima Taylor, Sai Krupa Das, Cheryl Gilhooly, Susan Roberts, and James Glass. 2019. A food logging system for iOS with natural spoken language meal descriptions (P21-009-19). *Current developments in nutrition* 3, Supplement_1 (2019), nzz041–P21. https://doi.org/10.1093/cdn/nzz041.P21-009-19

[42] Brian Y Lim, Xinni Chng, and Shengdong Zhao. 2017. Trade-off between automation and accuracy in mobile photo recognition food logging. In *Proceedings of the Fifth International Symposium of Chinese CHI*. 53–59. https://doi.org/10.1145/3080631.3080640

[43] Yuhan Luo, Bongshin Lee, and Eun Kyoung Choe. 2020. TandemTrack: shaping consistent exercise experience by complementing a mobile app with a smart speaker. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3313831.3376616

[44] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing food trackers with dietitians: Identifying design opportunities for food tracker customization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3290605.3300822

[45] Holger Lütters, Malte Friedrich-Freksa, and Marc Egger. 2018. Effects of speech assistance in online questionnaires. In *General Online Research Conference, Cologne, Germany*.

[46] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 451–462. https://doi.org/10.1145/2971648.2971677

[47] Elizabeth L Murnane, Dan Cosley, Pamara Chang, Shion Guha, Ellen Frank, Geri Gay, and Mark Matthews. 2016. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association* 23, 3 (2016), 477–484. https://doi.org/10.1093/jamia/ocv165

[48] Shuo Niu, D Scott McCrickard, Timothy L Stelter, Alan Dix, and G Don Taylor. 2019. Reorganize Your Blogs: Supporting Blog Re-visitation with Natural Language Processing and Visualization. *Multimodal Technologies and Interaction* 3, 4 (2019), 66. https://doi.org/10.3390/mti3040066

[49] Arnel B Ocay, Jane M Fernandez, and Thelma D Palaoag. 2017. NutriTrack: Android-based food recognition app for nutrition awareness. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*. IEEE, 2099–2104. https://doi.org/10.1109/CompComm.2017.8322907

[50] Hyungik Oh, Jonathan Nguyen, Soundarya Soundararajan, and Ramesh Jain. 2018. Multimodal food journaling. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*. 39–47. https://doi.org/10.1145/3264996.3265000

[51] Ronaldo Parente, Ned Kock, and John Sonsini. 2004. An analysis of the implementation and impact of speech-recognition technology in the healthcare sector. *Perspectives in Health Information Management* 1 (2004). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2047322/

[52] Somani Patnaik, Emma Brunskill, and William Thies. 2009. Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. In *2009 International Conference on Information and Communication Technologies and Development (ICTD)*. IEEE, 74–84. https://doi.org/10.1109/ICTD.2009.5426700

[53] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia van Gemert-Pijnen. 2014. Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing* 18, 7 (2014), 1667–1676. https://doi.org/10.1007/s00779-014-0779-y

[54] Margaret Raber, Tom Baranowski, Karla Crawford, Shreela V Sharma, Vanessa Schick, Christine Markham, Wenyan Jia, Mingui Sun, Emily Steinman, and Joya Chandra. 2020. The Healthy Cooking Index: Nutrition optimizing home food preparation practices across multiple data collection methods. *Journal of the Academy of Nutrition and Dietetics* (2020). https://doi.org/10.1016/j.jand.2020.01.008

[55] Melanie Revilla, Mick P Couper, Oriol J Bosch, and Marc Asensio. 2020. Testing the use of voice input in a smartphone web survey. *Social Science Computer Review* 38, 2 (2020), 207–224. https://doi.org/10.1177/0894439318810715

[56] Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James A Landay. 2018. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–23. https://doi.org/10.1145/3161187

[57] Samsung, Inc. 2021. Samsung Voice Recorder. Retrieved April 26, 2021 from https://www.samsung.com/au/support/mobile-devices/using-voice-recorder/

[58] Kshitij Saxena, Robert Diamond, Reid F Conant, Terri H Mitchell, Guido Gallopyn, and Kristin E Yakimow. 2018. Provider adoption of speech recognition and its impact on satisfaction, documentation quality, efficiency, and cost in an inpatient EHR. *AMIA Summits on Translational Science Proceedings* 2018 (2018), 186. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961784/

[59] Michael F Schober, Frederick G Conrad, Christopher Antoun, Patrick Ehlen, Stefanie Fail, Andrew L Hupp, Michael Johnston, Lucas Vickers, H Yanna Yan, and Chan Zhang. 2015. Precision and disclosure in text and voice interviews on smartphones. *PloS one* 10, 6 (2015), e0128337. https://doi.org/10.1371/journal.pone.0128337

[60] Benjamin Schüz, Natalie Schüz, and Stuart G Ferguson. 2015. It's the power of food: individual differences in food cue responsiveness and snacking in everyday life. *International Journal of Behavioral Nutrition and Physical Activity* 12, 1 (2015), 149. https://doi.org/10.1186/s12966-015-0312-3

[61] Lisa B Signorello, Heather M Munro, Maciej S Buchowski, David G Schlundt, Sarah S Cohen, Margaret K Hargreaves, and William J Blot. 2009. Estimating nutrient intake from a food frequency questionnaire: incorporating the elements of race and geographic region. *American journal of epidemiology* 170, 1 (2009), 104–111. https://doi.org/10.1093/aje/kwp098

[62] Lucas M Silva, Yuqi Huai, and Daniel A Epstein. 2020. Exploring Voice Input Opportunities in Multimodal Food Journaling. In *CHI'20 Workshop on Conversational Agents for Health and Wellbeing*.

[63] Rachel Silver, Mandy Korpusik, Salima Taylor, Sai Das, Cheryl Gilhooly, James Glass, and Susan Roberts. 2019. Testing the Validity of a Natural Spoken Language Application for the Self-Monitoring of Daily Dietary Intake (P13-035-19). *Current developments in nutrition* 3, Supplement_1 (2019), nzz036–P13. https://doi.org/10.1093/cdn/nzz036.P13-035-19

[64] Jolene D Smyth, Don A Dillman, Leah Melani Christian, and Mallory McBride. 2009. Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly* 73, 2 (2009), 325–337. https://doi.org/10.1093/poq/nfp029

[65] Jeffery Sobal and Carole A Bisogni. 2009. Constructing food choice decisions. *Annals of Behavioral Medicine* 38, suppl_1 (2009), s37–s46. https://doi.org/10.1007/s12160-009-9124-5

[66] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3988–3999.

[67] TeamViewer. 2021. TeamViewer QuickSupport. Retrieved April 26, 2021 from https://www.teamviewer.com/en-us/info/quicksupport/

[68] Nađa Terzimehić, Christina Schneegass, and Heinrich Hussmann. 2018. Towards finding windows of opportunity for ubiquitous healthy eating interventions. In *International Conference on Persuasive Technology*. Springer, 99–112. https://doi.org/10.1007/978-3-319-78978-1_8

[69] WA Van Staveren, JO De Boer, and J Burema. 1985. Validity and reproducibility of a dietary history method estimating the usual food intake during one month. *The American journal of clinical nutrition* 42, 3 (1985), 554–559. https://doi.org/10.1093/ajcn/42.3.554

[70] Stephen H Walsh. 2004. The clinician's perspective on electronic health records and how they can affect patient care. *Bmj* 328, 7449 (2004), 1184–1187. https://doi.org/10.1136/bmj.328.7449.1184

[71] Yunlong Wang, Ulrike Pfeil, and Harald Reiterer. 2016. Supporting self-assembly: The IKEA effect on mobile health persuasive technology. In *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*. 19–22. https://doi.org/10.1145/2985766.2985775

[72] Matthias Wenzel, Anja Perlich, Julia PA von Thienen, and Christoph Meinel. 2019. New ways of data entry in doctor-patient encounters. In *Design Thinking Research*. Springer, 159–177. https://doi.org/10.1007/978-3-319-97082-0_9

[73] Yixuan Zhang and Andrea G Parker. 2020. Eat4Thought: A Design of Food Journaling. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8. https://doi.org/10.1145/3334480.3383044

[74] Zoom Video Communications, Inc. 2021. *Video Conferencing, Web Conferencing, Webinars, Screen Sharing - Zoom.* Retrieved April 26, 2021 from https://zoom.us/