

A Contextual Inquiry of People with Vision Impairments in Cooking

Franklin Mingzhe Li
Carnegie Mellon University
Pittsburgh, PA, United States
mingzhe2@cs.cmu.edu

Shaun K. Kane
Google Research
Boulder, Colorado, United States
shaunkane@google.com

Michael Xieyang Liu
Google Research
Mountain View, CA, United States
lxieyang@google.com

Patrick Carrington
Carnegie Mellon University
Pittsburgh, PA, United States
pcarrington@cmu.edu

ABSTRACT

Individuals with vision impairments employ a variety of strategies for object identification, such as pans or soy sauce, in the culinary process. In addition, they often rely on contextual details about objects, such as location, orientation, and current status, to autonomously execute cooking activities. To understand how people with vision impairments collect and use the contextual information of objects while cooking, we conducted a contextual inquiry study with 12 participants in their own kitchens. This research aims to analyze object interaction dynamics in culinary practices to enhance assistive vision technologies for visually impaired cooks. We outline eight different types of contextual information and the strategies that blind cooks currently use to access the information while preparing meals. Further, we discuss preferences for communicating contextual information about kitchen objects as well as considerations for the deployment of AI-powered assistive technologies.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in accessibility**.

KEYWORDS

Cooking, Contextual Inquiry, Blind, People with Vision Impairments, Accessibility, Assistive technology

ACM Reference Format:

Franklin Mingzhe Li, Michael Xieyang Liu, Shaun K Kane, and Patrick Carrington. 2024. A Contextual Inquiry of People with Vision Impairments in Cooking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3613904.3642233>



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642233>

1 INTRODUCTION

Cooking holds a profound sway over the overall quality of life for individuals with vision impairments [5, 39]. Nevertheless, the culinary process leans heavily on visual cues, relying extensively on the contextual information of objects nestled within the kitchen (e.g., location, status), which remains elusive to people with vision impairments [39, 60]. This creates substantial barriers for them to cook independently and, ultimately, negatively impacts their quality of life [27]. For example, seemingly simple tasks like finding an item in a cluttered refrigerator or gauging whether a dish is fully cooked are challenging for visually impaired individuals, particularly for beginners in rehabilitation training programs [39, 63]. Prior research has highlighted the importance of recognizing contextual information [60], such as the color and placement of objects in everyday tasks. However, there exists a notable gap of knowledge regarding the specific contextual cues that visually impaired people rely on and the rationale behind these preferences when cooking.

Assistive vision systems like SeeingAI [47] and TapTapSee [58] have proven effective in helping visually impaired people identify objects under various settings. However, these systems frequently fall short in the kitchen due to usability challenges and impractical designs [39]. For instance, past studies have highlighted their lackluster performance in recognizing clusters of objects or retrieving precise information about an object, such as the expiration date of products [1, 14]. Additionally, these systems often require holding a smartphone to capture information, which is inconvenient and impractical during cooking tasks that are time-sensitive [39]. Therefore, it is crucial to understand how visually impaired individuals obtain contextual information about kitchen objects as well as explore effective ways to communicate this information while cooking, thereby shedding light on opportunities to increase culinary independence and enhance their overall quality of life.

To summarize, in this work, we aim to investigate:

- RQ1 What contextual information about objects is important to people with vision impairments during cooking?
- RQ2 What techniques and strategies do blind cooks use to non-visually gather and utilize information about objects during cooking and food preparation tasks?
- RQ3 What are the most effective methods for conveying necessary contextual information about objects at the right times and in the right ways?

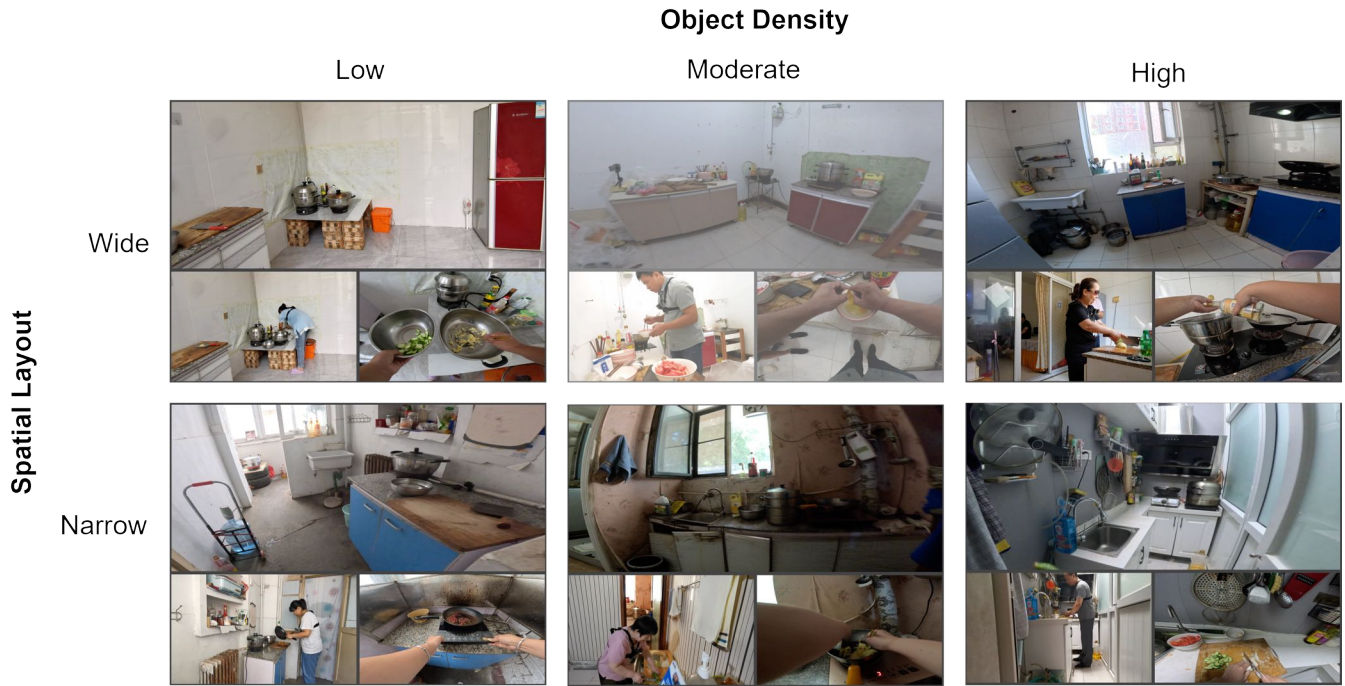


Figure 1: Contextual Inquiry: sample views of participants' home kitchens with varying spatial layouts (from wide to narrow), as well as object density.

To address our research questions, we conducted a comprehensive contextual inquiry study with 12 visually impaired individuals experienced in cooking (Figure 1). This inquiry took place in participants' own kitchens, where they were asked to naturally prepared meals using ingredients available in their refrigerators (as detailed in Section 3.2.2). Subsequently, they were instructed to perform specific kitchen-related tasks related to object recognition, such as identifying ingredients and organizing their kitchen space). In the end, we conducted semi-structured interviews to gain further insights into the motivations and rationales behind their actions, understand their needs for object-related information, and explore their views on integrating AI-powered assistive technologies in the kitchen (Section 3.2.3).

Our study offers a comprehensive analysis of the contextual information requirements for objects, comprising five primary, two secondary, and one application-specific category (Section 4). We explore the nuanced process of gathering contextual information about objects for cooking activities, which involves establishing intentional contextual associations with objects (Section 5). Furthermore, we examine factors related to displaying and communicating contextual information, including adjustable information verbosity (Section 6). Finally, to provide a holistic perspective, we discuss the design implications for AI-powered assistive technologies tailored for visually impaired individuals in kitchen environments (Section 7).

Our work makes the following contributions:

- A **contextual inquiry study** with 12 visually impaired individuals experienced in cooking, conducted in their own

kitchens, yielding novel insights into their challenges and needs regarding objects' contextual information.

- A **taxonomy of objects' contextual information need** for cooking, including primary, secondary, and application-specific information, which fills a notable gap in existing research and offers a systematic guide for developing future systems to support people with vision impairments in the kitchen.
- A documentary of the **existing process of obtaining contextual information non-visually and associated challenges**, highlighting how visually impaired cooks create intentional associations with objects, enriching our understanding of their cooking experiences.
- A discussion of **the strategies for presenting and communicating contextual information** with future AI-powered assistive technologies.

2 RELATED WORK

In this section, we first present the background knowledge of cooking experiences by people with vision impairments. We then describe the related work of information identification for people with vision impairments, followed by existing AI-powered technologies for kitchen space.

2.1 Cooking Experiences by People with Vision Impairments

Vision impairments have been shown to significantly affect individuals' experiences related to food, eating, and cooking [5]. In a

study involving over 100 visually impaired individuals, Jones et al. discovered a substantial correlation between the severity of vision impairment and the challenges faced in shopping for ingredients and preparing meals [27]. This research also highlighted a concerning connection between vision impairments and malnourishment, ultimately leading to a diminished quality of life [27]. Bilyk et al. [5] conducted semi-structured interviews with nine visually impaired individuals, revealing their heavy reliance on prepared food from external sources. Notably, all participants in the study reported consuming a minimum of 40% of their dinners in restaurants to avoid the challenges associated with cooking [5]. More recent findings from Kostyra et al. [34] indicated that out of 250 survey respondents, 49.6% of visually impaired individuals prepare their meals independently, while others seek assistance from sighted and/or blind individuals. Notable difficulties in food preparation included peeling vegetables (82.1% reported difficulty) and frying foods (72% reported difficulty). Conversely, tasks not requiring heat or specialized tools, such as preparing sandwiches and washing fruits, were reported as more manageable. Given the ease of preparation, 57.6% of participants opted for ready-to-eat products, while only 14% preferred ready-to-heat meals [34].

Prior research has highlighted different practices and challenges that people with vision impairments have in cooking activities [5, 39, 63]. For example, Li et al. [39] uncovered various difficulties for people with vision impairments while cooking, such as measuring, organizing space, tracking objects, and quality inspection. Among the tasks described by people with vision impairments in the kitchen [5, 39, 63], such as tracking objects and organizing space, many are relevant to the identification of different contextual information of objects (e.g., shape, color, location). Despite these existing explorations into the cooking experiences of people with vision impairments, there remain unclear regarding what contextual information of objects is needed across different cooking procedures.

2.2 Information Identification for People with Vision Impairments

Cooking often requires people with vision impairments to obtain contextual information about objects while cooking. Prior research has explored opportunities of adding tactile markers to devices or objects [23, 57]. For example, Guo et al. [23] created 3D printed tactile marking to better support people with vision impairments to interact with different interfaces. Beyond adding tactile markers, prior work also explored using crowdsourcing [4, 24] or computer vision [4, 18, 22, 30, 48, 59] to identify objects of interest. For example, Vizwiz [4] introduced a crowdsourcing-based approach for mobile phones that answers visual questions in nearly real-time, such as the color of objects. Moreover, VizLens leveraged computer vision and crowdsourcing to enable people with visual impairments to interact with different interfaces, such as a microwave oven [22]. Beyond supporting object recognition, Zhao et al. [65] also explored how should visual information be presented to people with vision impairments and how should such system guide people to the targeted objects. Given existing approaches to identifying contextual information of objects by people with vision impairments, little has been explored to understand what are the contextual information

needs in cooking scenarios, as well as how such systems should be developed to support cooking-related tasks.

2.3 AI-Powered Technology for Activities of Daily Living in the Kitchen

AI-based kitchen technologies have been widely explored in HCI to enhance people's quality of life, such as monitoring kitchen activities and objects [36, 49], supporting multimodal control and automation with kitchen appliances [6, 31, 62], enabling sensing capabilities for smart tools and utensils [33], and providing dynamic guidance for cooking instructions [9, 11, 17, 35]. For example, Lei et al. [36] deployed an RGB-D camera inside the kitchen space and used RGB-D cameras to recognize fine-grained activities that include both activity and object recognition. Furthermore, Konig and Thongpull [33] invented Lab-on-Spoon, a 3D integrated multi-sensor spoon system for detecting food quality and safety, such as temperature, color, and pH value, to differentiate ingredients like fresh oil vs. used oil. Beyond recognizing activities and objects with specific technologies, prior research also stated the importance of providing full experiences for people in the kitchen space with deployments of both hardware and software [6]. Although prior research has explored different AI-powered applications for kitchen activities it is unknown what contextual information is important to people with vision impairments in the kitchen and how systems such as these might need to be adapted for people with vision impairments.

3 CONTEXTUAL INQUIRY STUDY OF BLIND COOKING

Contextual inquiry has been used in field research to understand end users' experiences, preferences, and challenges during their everyday activities [3, 26, 28, 43, 44, 52]. Contextual inquiry research can support deeper understandings of everyday human behaviors through surfacing facts, details, constraints, and structures [3]. Contextual inquiry has been widely adopted as a research method to better understand marginalized groups, cultural learning, and accessibility practices (e.g., [12, 13, 46, 55]). To investigate our research inquiries comprehensively, we initiated a **contextual inquiry involving 12 participants with vision impairments naturally cooking in their kitchens**. Our study unfolds in three distinct phases: the pre-study interview, the contextual inquiry, and the semi-structured interview.

3.1 Participants

We recruited 12 people with vision impairments from the mailing list of the China Disabled Persons' Federation (Table 1). To participate in our study, participants were required to be 18 years or older, legally or totally blind, and had prior experience with cooking. Among the 12 participants we recruited, six of them are female and six are male (Table 1). The average age of our participants was 42.5 (SD = 6.1). Eight of them are totally blind and four are legally blind (Table 1). Regarding the four participants who are legally blind, P1 has no vision in the left eye and light perception in the right eye. P3, P8, and P12 have some light perception in both eyes. They had an average of 21.7 years of cooking experience (SD = 9.0). As per their self-reports, seven of them cook every day, one cooks three



Figure 2: Contextual inquiry settings. Three camera views captured during the study to show the overall kitchen environment, ego-centric view, and a 3rd-person camera view of the person cooking. The top image is the overall still image of the kitchen setting. The bottom left image is the view from the stationary camera that shows our participant making a meal. The bottom right image is the egocentric view captured from the chest-mounted camera.

or four times a week, two of them cook once per week, one cooks two or three times a month, and one cooks once per month (Table 1). Regarding living arrangements, eight resided with their families, one with roommates, and the remainder lived independently. (Table 1). Participants were compensated in local currency which is equivalent to \$40 USD. The recruitment and study procedure was approved by our organization’s Institutional Review Board (IRB). Each participant’s engagement took approximately 90 to 105 minutes.

3.2 Study Procedure

Our study unfolds in three distinct phases: a pre-study interview, a contextual inquiry, and a semi-structured interview.

3.2.1 Pre-study Survey [5 Minutes]. In the initial pre-study interview, we gathered demographic information from our participants. This included details such as age, gender, vision condition, cooking experience, cooking frequency, living arrangements, preferred cooking activities, and any challenges they encountered in the culinary domain.

3.2.2 Contextual Inquiry [75 Minutes]. During the contextual inquiry phase, participants engaged in various cooking activities within their personal kitchens (as depicted in Figure 2). To record their culinary experiences comprehensively, participants were equipped with a GoPro 11 camera [20] attached to their chests, allowing us to document their actions and behaviors (as shown in Figure 2). To provide a comprehensive view of their activities, we installed a

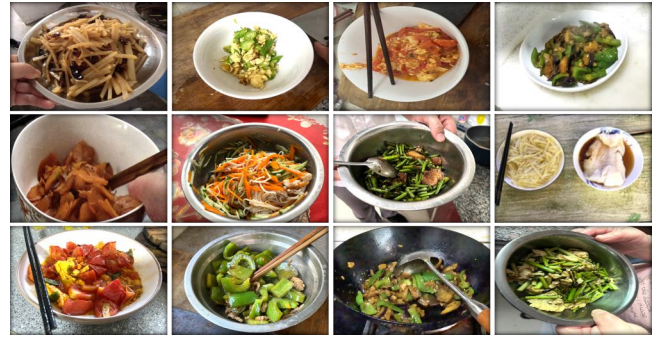


Figure 3: Dishes made by our participants during the contextual inquiry in their kitchens.

stationary camera within their kitchen environments (Figure 2). In the contextual inquiry, there were two main tasks:

[TASK 1: Self-Directed Cooking]: To observe a full experience of cooking, our participants were asked to first explore the food and ingredients that they have in their own kitchen, and then make a dish based on the availability of ingredients (See Figure 3 for dishes made by our participants). After making the dish, participants were asked to serve the dish and clean the kitchen. During the cooking process, participants were encouraged to vocalize their thoughts while performing tasks, following the think-aloud protocols throughout the contextual inquiry [61].

[TASK 2: Specific Cooking-related Activities]: Following the completion of Task 1, participants were further engaged in specific cooking-related activities to gain a deeper insight into their procedural methods and information requirements. Drawing from prior research findings [39], we chose eight key activities that epitomize the essential cooking tasks for individuals with vision impairments that are relevant to object identification. These activities encompassed: 1) identification of ingredients and food items, 2) recognition of cookware and utensils, 3) precise measurement, 4) monitoring of cooking progress, 5) the process of serving food, 6) ensuring safety measures, 7) maintaining kitchen organization, and 8) executing grocery shopping. Participants were requested to demonstrate their typical approach to completing these tasks and to vocalize their thought processes throughout, adhering to the think-aloud protocols [61].

3.2.3 Semi-structured Interview [25 Minutes]. Following the completion of the contextual inquiry, we proceeded to conduct a structured semi-structured interview. This interview served as a platform for a comprehensive debriefing, allowing us to delve into our observations and findings from the contextual inquiry. Additionally, it provided an opportunity to explore various facets of future design considerations related to the communication of visual information and the form factors of assistive technologies within the kitchen environment.

[Debriefing of the Contextual Inquiry]: In this segment of the interview, we initiated discussions concerning the behaviors and processes observed during the contextual inquiry (refer to Section 3.2.2). We delved deeper into the challenges encountered by participants, as well as their specific visual information needs. This

PID	Age	Gender	Vision Condition	Cooking Experiences and Frequency	Living Condition
P1	41	Male	Legally Blind	16 Years, Three or Four Times a Week	Living with Family
P2	40	Male	Totally Blind	18 Years, Once per Month	Living with Family
P3	32	Female	Legally Blind	12 Years, Everyday	Living with Family
P4	46	Male	Totally Blind	24 Years, Once per Week	Living with Family
P5	52	Male	Totally Blind	29 Years, Once per Week	Living with Family
P6	44	Female	Totally Blind	15 Years, Everyday	Living with Family
P7	49	Male	Totally Blind	26 Years, Everyday	Living Independently
P8	36	Female	Legally Blind	10 Years, Everyday	Living with Family
P9	39	Male	Totally Blind	18 Years, Everyday	Living Independently
P10	42	Female	Totally Blind	36 Years, Everyday	Living Independently
P11	51	Female	Totally Blind	38 Years, Everyday	Living with Friends
P12	38	Female	Legally Blind	18 Years, Two or Three Times a Month	Living with Family

Table 1: Demographic information of our study participants

encompassed topics such as object localization, contextual information regarding objects, and the dynamics of visual information requirements throughout different stages of the cooking process.

[Design of Information Communication and Deployment]: Subsequently, our conversation shifted towards the design aspects related to information communication and deployment within the kitchen setting. We inquired about participants' preferences for how information should be conveyed to them while cooking, their information needs during culinary activities, their opinions on the form factors of assistive technologies within the kitchen, and any other concerns or considerations they wished to share.

3.3 Data Analysis

The contextual inquiries were meticulously documented using both video recording methods. High-resolution footage was captured at 5.3k resolution with a frame rate of 30Hz utilizing the GoPro 11 camera in HyperView [20]. Additionally, a stationary camera was employed to supplement the recordings. Meanwhile, the semi-structured interviews were recorded in audio format. We leveraged the video resources captured by the body-worn camera and the stationary camera for the **contextual inquiry** analysis [32]. For the analysis of the contextual inquiry videos, we used thematic analysis [8]. Two researchers independently annotated the video and open-coded the observations. Our analysis centered on two key aspects: the contextual information needs related to objects and the processes involved in acquiring contextual information about these objects (as detailed in Section 3.2.2). As for the analysis of the **semi-structured interviews**, a similar thematic analysis approach was employed [8]. This analysis revolved around themes related to the processes of acquiring contextual information, expected methods of information communication for objects, visual information requirements, and considerations regarding form factors (refer to Section 3.2.3).

4 FINDINGS: CONTEXTUAL INFORMATION NEEDS OF OBJECTS (RQ1)

In this section, we first present the five fundamental contextual information needs of objects that people with vision impairments prefer while cooking: position information, orientation information, proximity and grouping information, similarity and duplicate information, and internal state information. We further present three secondary and application-specific information needs that are relevant to kitchen activities: safety-related information, health-related information, and plating and serving information. We illustrated this in the Table 2.

4.1 Five Fundamental Categories of Contextual Information for Objects

4.1.1 Position Information. During our contextual inquiry conducted in the participants' kitchens, we uncovered a unanimous consensus among our participants regarding the paramount importance of knowing the precise position of objects within their culinary domains, which could reduce the time effort of finding objects while cooking (9), and support autonomy and agency in the kitchen (3). We found over half of our participants encountered difficulties when attempting to locate specific items while cooking, due to the mental load of memorizing the positions (P1, P8, P11) or misplacement of items by their family members or friends (P2, P4, P6, P10). This challenge increased when dealing with less frequently used items. To remember object locations, our participants typically advocated for the **use of a reference point in the kitchen, or an "anchor," to indicate the object's position**. For instance, they preferred describing the oyster sauce as being "on the windowsill" rather than specifying its coordinates in 3D space (e.g., x, y, z). P5 elucidated:

"To determine the location of the item I'm searching for, I simply require a relative position in relation to

Attribute	Description	Priority
Position	The relative location to a reference point, or an “anchor,” to indicate the object’s position	Primary
Orientation	Information about how an object is currently oriented (e.g., vertical, horizontal, or tilted) relative to a reference point (e.g., human, object)	Primary
Proximity and Grouping	Information about groupings of objects relative to others and the environment	Primary
Similarity and Duplicates	Information about similar or duplicate objects, which includes differentiation between similar objects, relative positions between objects that are similar, and the overall quantity of similar objects	Primary
Internal State	Information about internal state of objects, such as cleanness, freshness, boiling water, the amount of solid or liquid inside the container, and the doneness of food	Primary
Safety-related Information	Information to monitor anything that might be harmful (e.g., knock over objects, flaming)	Secondary
Health-related Information	Information to track objects with health hazards after consumption (e.g., expired food, overcooked food)	Secondary
Plating and Serving	Information about the final appearance and presentation of objects upon finishing (e.g., color distribution)	Application-Specific

Table 2: Contextual information needs revealed during our study. Primary attributes were required to interact with objects, while secondary and application-specific attributes were related to understanding the state of objects and manipulating them.

a reference point in my kitchen, such as near my gas range or on my fridge.”

4.1.2 Orientation Information. Our inquiry also shed light on another critical facet - the need to ascertain the orientation of various kitchen objects, signifying their **current alignment or positioning relative to a specific reference point or another object**. We found that our participants encountered challenges in discerning the orientation of objects during different culinary tasks (P2, P4, P5, P6, P9, P11). For example, both P2 and P5 grappled with slicing pork belly with the correct orientation, often leading to the unintended separation of fat and lean meat portions. Similarly, our participants faced recurrent tribulations when endeavoring to align a wok precisely with plates or bowls during the serving process (P4, P6, P9, P11). This issue invariably resulted in unwanted food spillage. P11 provided illuminating insights into the complexities of this matter:

“Perfectly aligning the wok with the plate during serving can be quite challenging, and as a consequence, I frequently find myself grappling with food spillage, necessitating subsequent cleanup efforts.”

4.1.3 Proximity and Grouping Information. In addition to object orientation, our participants placed significant emphasis on contextual information related to proximity and grouping - essentially, the **arrangement and organization of objects in relation to one another and the kitchen environment**. Knowing the proximity and grouping information of certain object groups can support

maintaining the kitchen, as well as navigating the kitchen space. This aspect encompassed the pressing need to determine if objects were correctly placed, if any misplacements or disarray had occurred (P4, P11), and whether alterations or rearrangements to the kitchen space had transpired during or after culinary activities (P5, P9). For instance, we observed that P5 faced difficulties locating sauces and ingredients following his daughter’s cooking session, as she had inadvertently rearranged various items. P5 articulated his predicament:

“I usually...have my sauces like soy sauce and vinegar arranged on the second shelf and other solid ingredients such as sugar and salt carefully positioned on the third shelf. However, my daughter cooked a meal yesterday, and the displacement of items made it exceedingly challenging to locate things today.”

4.1.4 Similarity and Duplicates Information. Our research also illuminated the critical need to acquire **information about similar or duplicate objects**, encompassing differentiation between similar items, determining the relative positions of objects, and assessing the overall quantity of similar objects. Understanding the quantity of specific items, such as tomatoes within the fridge, emerges as vital for monitoring food supplies, particularly when preparing for grocery shopping. During our study, we observed that P3 had multiple tomatoes located at various spots within her refrigerator. When asked about the tomatoes’ whereabouts, she discovered two with holes and remarked:

“I wasn’t aware that I had these tomatoes tucked away in the corner of my fridge, and I can’t recall how long they’ve been there. It’s possible my son placed them there.”

Moreover, we noted that providing information regarding the differentiation between similar objects can significantly assist individuals with vision impairments in comprehending their kitchen environments. For example, P1 and P8 both expressed the desire to know how different plates were stacked together on a shelf and whether all the plates were identical or if any variations existed during the serving process.

4.1.5 Internal State Information. Kitchen objects may possess various **internal states, including temperature, freshness, cleanliness, the condition of solid or liquid contents, and the degree of doneness for food that is being cooked.** Our participants noted that ascertaining these internal states can be a challenge, primarily because many of these assessments rely on visual cues. For instance, our participants expressed difficulties in determining the cleanliness of vegetables or meat (P2, P9). Other challenging tasks include monitoring the water temperature to determine if it has reached boiling point (P5) or gauging the readiness of food (P6). P9 explained this issue:

“It’s impossible for me to determine if vegetables are clean or not during the washing process. Consequently, I often find myself repeatedly washing them to ensure their absolute cleanliness.”

4.2 Secondary and Application-specific Information

In addition to the five fundamental contextual information categories, we identified three types of information that people with vision impairments are acutely aware of while interacting with objects in the kitchen.

4.2.1 Safety-related Information. The first secondary information is safety-related information, which encompasses monitoring objects that have the potential to cause harm. This includes objects that could be accidentally knocked over during kitchen tasks (P4, P9, P10) and safety-related concerns, such as monitoring the temperature of cooking equipment to prevent accidents (P4, P5, P6, P10, P11). Our observations revealed instances where P4 accidentally knocked over a salt bottle while searching for a plate, and P9 tipped over a nearby water cup. P9 emphasized the significance of being aware of potential obstacles:

“Being aware of potential obstructions that I might knock over could greatly benefit me throughout the preparation and cooking process. It could also help reduce the anxiety associated with interacting with objects while learning.”

4.2.2 Health-related Information. Our contextual inquiry highlighted the critical importance of knowing contextual information about objects for health-related considerations. This encompasses factors like checking expiration dates (P1, P10, P12), identifying food with potential health risks, such as overcooked items (P2, P5), and recognizing visual cues on food items, such as stickers on



Figure 4: Some visual information can have health implications. Left: Sliced tomatoes still have stickers attached (highlighted); Right: Some vegetables were overlooked and left on the cutting board (highlighted).

vegetables (P6, P8). P2 articulated the challenges he faced in this regard:

“To ensure all of the meat is fully cooked, I usually cook it for a longer period, which sometimes had some of the food got overcooked or even burned.”

We also observed instances where participants unintentionally left behind or missed certain food items, especially when dealing with round-shaped objects like green beans, or during the transfer of items from the cooking vessel to the plate. This oversight could lead to health-related concerns, such as consuming spoiled food or creating conditions favorable to pests like cockroaches (P2, P4, P5, P6, P11) (Figure 4). For example, P5 inadvertently included a food sticker while slicing a tomato, and it was subsequently cooked in the dish (Figure 4). P5 expressed:

“I had no idea there were stickers on the tomato, and it’s challenging to use my hands to feel the entire tomato to detect the sticker. Wouldn’t it be better if they stopped using stickers altogether?”

4.2.3 Plating and Serving Information. Upon completing the cooking process, we uncovered our participants’ preferences for being informed about the presentation and final appearance of the dishes they prepared. This encompassed details such as the arrangement of items on the plate (position), the degree of cooking (internal state), the spatial relationship with other items on the plate (proximity), and the orientation of food items (orientation). P5 emphasized the value of being aware of the visual presentation of their culinary creations:

“It’s valuable for me to be aware of how appealing my food looks once it’s served, or if I should consider adding additional vegetables or meat to enhance the overall visual presentation.”

5 FINDINGS: PROCESS OF OBTAINING CONTEXTUAL INFORMATION OF OBJECTS FOR COOKING ACTIVITIES (RQ2)

In this section, we begin by discussing how individuals with vision impairments rely on multiple sensory inputs to acquire various types of contextual information [indicated in brackets] concerning objects during cooking activities. These sensory inputs encompass touch, sound, and smell. Subsequently, we explore two distinct approaches employed by our participants to streamline the process



Figure 5: P9 using his left hand to touch the side of the wok to gauge the temperature of the wok.

of identifying objects of interest: the creation of supplementary contextual information and the simplification of the contextual information acquisition process related to objects.

5.1 Multi-sensory Approaches to Obtaining Contextual Information

5.1.1 Touch. Our contextual inquiry revealed that our participants relied on their sense of touch for tracking the status of objects in their kitchen, including dual-purpose scanning and memorization, manipulating objects, and performing safety inspections.

Dual-purpose Scanning and Memorization: We found that our participants used touch for multiple purposes, such as localizing objects and checking the internal state of objects. This included locating items like vegetables, fruits, and meat within the refrigerator, as well as identifying sauces, ingredients, or containers within the kitchen [**Position**]. Touching food also communicated information about the freshness of the food [**Internal State**]. P4 commented: *“Using touch is my primary way of finding things in the kitchen, while I explore objects, I also feel the object to know if it is fresh through the stiffness or if there are holes on the skin.”*

During this process, they engaged in pre-organizing objects of interest and often memorized the positions of other similar or nearby objects as they scanned the space [**Proximity and Grouping**][**Similarity and Duplicates**]. For instance, P1 searched for eggplants in the fridge, scanning through it while simultaneously committing the location of garlic to memory. Later, when he needed minced garlic, he easily found it, remarking, *“I memorized the position of the garlic last time when I was scanning through the fridge!”* However, we found this approach sometimes can take a long time to scan through objects, and people might miss certain objects through scanning due to the complexity of the space and form a wrong memory of objects and space (P3, P7, P11). P3 explained:

“It often takes me a while to find the vegetable that I want to get. And it is easy for me to miss some of it, because kitchen shelves and refrigerator storage are complex, such as my tomatoes were placed at multiple positions in the kitchen. Once I did not find it, then it might just stay in a corner for many days.”

Precision and Manipulation: While manipulating and interacting with objects, we found our participants also leverage touch to ensure objects are organized and aligned, which maintained order and reduced the risk of spillage, [**Orientation**] as well as to count and measure objects (e.g., sugar, vinegar) [**Internal State**]. To align objects, or when transferring materials between containers (e.g., adding sauces to salad, serving food from a wok) our participants usually used one hand to hold the object then used another hand to find and secure the other object. To determine the quantity of dry ingredients such as salt and sugar, they relied on touch, using their hands to feel and specify the exact amount (P1, P4, P5, P8, P9). For liquids, they often placed a finger beneath the lid, allowing them to feel the liquid passing through their finger to gauge the quantity (P5, P7, P10).

Safety Inspection: Furthermore, we discovered that access to safety-related information could significantly reduce the risks associated with cooking for individuals with vision impairments. For example, both P9 and P10 routinely performed thorough inspections of flammable objects that are close to the gas range before cooking [**Proximity and Grouping**][**Safety**]. P10 expressed a desire for pre-cooking safety checks: *“I would appreciate having some form of support for conducting safety checks before cooking to ensure there are no objects in close proximity to the range during cooking.”* Additionally, we found that gauging the temperature of the wok [**Internal State**][**Safety**] presented one of the most formidable challenges for individuals with vision impairments, as this information was traditionally obtained through tactile means, such as direct touch (see Figure 5). This practice, although effective, often resulted in burns and blisters, as expressed by P9:

“I use my hand to feel the temperature of the wok; you can see my arm has many blisters and burns, but I have to use this method as there is no other way for me to gauge the temperature or balance the wok correctly.”

5.1.2 Sound. Recognizing and tracking sounds played a pivotal role in our participants’ ability to assess the status of objects [**Internal State**] in cooking, such as temperature. As an illustrative example, P11 described a method involving the addition of a small amount of egg yolk to hot oil in a wok to listen to the resulting sound to estimate the oil’s temperature. While sounds were usually helpful, it was sometimes difficult to follow sounds because of excessive background noise (9), such as kitchen exhaust fans or conversations with others in the room. P11 provided further insights into this issue:

“I prefer using sound-based assessments, such as checking if water has boiled, but sometimes the differences in sound characteristics can be quite subtle...which is difficult to specify with my exhaust fan on.”

In addition to assessing object status, participants utilized sound to estimate the quantity or volume of objects [**Internal State**]. An interesting example involved the use of containers with narrow nozzles. When pouring liquids into a pan or wok, the air inside the container compresses as the liquid flows, generating a distinct sound, often described as a “burp.” This auditory cue allowed individuals to approximate the amount of liquid dispensed. P8 detailed



Figure 6: Unique containers differentiate contents. Participants use and reuse unique containers to help identify objects.

the practicality of this method for estimating quantities through sound while acknowledging its inherent limitations, such as reduced precision, particularly when the liquid level in the container was low: *“I use sound to estimate how much oil I’ve poured into the pan. It’s challenging to discern through touch alone. The first drop of oil hitting the wok generates a small sound, and since the oil bottle I purchased has only one nozzle, it doesn’t continuously pour. Instead, it dispenses intermittently, producing a ‘burp’ sound. I rely on this auditory feedback to gauge the amount of oil in the pan. However, it’s not always precise, as the sound may become less noticeable when the liquid level in the bottle is low.”*

5.1.3 Smell. The sense of smell plays a critical role in helping individuals with vision impairments determine the internal state of objects and health-related information [Internal State][Health], such as the freshness or the doneness of the food. In cases where touch exploration was not feasible or practical, participants relied on their olfactory senses to detect signs of spoilage. P3 illustrated this practice, stating, *“It is common for us to have leftovers of main dishes as well as rice and bread. Sometimes, touching is often not feasible to assess the condition of the food. So, I use my sense of smell to determine if the food has gone bad. Spoiled food often emits a sharp and unpleasant smell due to fermentation.”* This reliance on smell allowed them to make informed decisions about whether it was safe to consume leftover food items.

Participants also utilized their sense of smell as a means to determine the readiness of certain dishes [Internal State]. Specific foods emitted distinctive aromas when they were close to being fully cooked. For example, P4 and P9 mentioned that particular dishes, such as those containing green peppers and meat, would release a savory aroma, signaling that they were nearly done. This olfactory cue served as an indicator of cooking progress. However, participants emphasized the importance of swift action once these aromatic cues were detected, as there was little room for delay between sensing the enticing aroma and preventing the food from becoming overcooked or burned. As P4 humorously put it, *“I like to use smell to gauge the readiness of my food. When you catch that aroma, it feels like a culinary achievement. But don’t celebrate too long; you need to promptly remove the food from the wok to prevent it from burning.”*



Figure 7: P4 used a knife as a “container” for scallions or minced garlic.

5.2 Altering Objects to Ease Identification

In addition to their multi-sensory strategies, our participants demonstrated the approach of actively creating additional contextual information for objects, thereby facilitating their recognition and organization. Through contextual inquiry, we uncovered a pervasive practice among all participants: the deliberate customization of objects to imbue them with supplementary contextual information. This approach involved introducing distinctive attributes, such as unique container shapes or deliberate organizational strategies, to enhance object identification and ensure secure organization within the kitchen.

5.2.1 Using Containers with Unique Shapes. We found that all of our participants typically chose to use different sizes or shapes of containers to indicate the difference between sauces, oil, or seasonings [Similarity and Duplicates] (Figure 6), which correspond to prior research showed that bartenders used different glasses to remember orders [16]. From the observation of the contextual inquiry, we found that they easily spotted the sauce that they wanted, and we found that for P1, he used a thin-headed, mid-sized glass to store oyster sauce, a large, rectangle bottle to store vinegar, and a wide-headed, large plastic bottle to store soy sauce. P5 also refills oils to the bottle that he purchased a long time ago (Figure 6). P6 further explained:

“I personally use different shapes of containers to indicate the difference of oyster sauce, vinegar, and soy sauce. There was one time that my daughter purchased a new bottle of oyster sauce, which confused me for many days.”

5.2.2 Securing Objects for Organization. Participants favored using bowls or plates to secure objects, promoting better organization, rather than letting them roll freely on the counter [Orientation][Position]. For example, P4, P5, P11, and P12 placed cut vegetables inside a bowl for ease of management. Similarly, P4, P6, and P10 employed a knife as a makeshift “container” to secure objects before cooking, such as scallions or minced garlic (Figure 7). This additional layer of context minimized the risk of accidents and lightened the cognitive load associated with memorizing object locations.

5.2.3 Spatial Separation and Grouping of Objects. In addition to enhancing the identification of individual objects, our study revealed

that participants deliberately arranged objects in separate spatial groupings based on their purposes [**proximity and grouping**], aiming to facilitate easier identification and access (P2, P3, P6, P8, P9, P11). During interviews, participants elaborated on the advantages of spatially separating objects, highlighting how it added another layer of context and improved the spatial recognition of object groupings. P8 provided insights into this behavior:

“I prefer to keep my kitchen essentials minimal and uncomplicated. Simultaneously, I adopt a practice of spatially separating different objects, aiding me in distinguishing between various items more effectively and reducing the chances of encountering unwanted objects due to spatial constraints.”

Furthermore, participants mentioned their practice of organizing objects within dedicated mini-spaces, each containing a specific group of items. For instance, P11 stored all sauces in the same location, while P9 demonstrated how he arranged daily-use bowls within easily accessible spaces and placed less frequently used ones in cabinets:

“Locating and reaching certain spaces can be quite demanding, especially when I have to stoop down to access bowls near ground level. Therefore, I keep utensils I use daily on the counter, within easy reach, and store the others in cabinets.”

5.3 Optimizing Cooking Procedures

In their pursuit of greater convenience and efficiency, our participants devised a range of strategies to simplify their cooking procedures, reducing the need for extensive contextual information gathering. These innovative methods encompassed pre-assigning orders, sequential organization, strategic seasoning placement according to usage frequency, and preserving spatial arrangements during the cleaning process. Through these tactical approaches, they were able to optimize their cooking routines while minimizing the effort required to access contextual information about objects.

5.3.1 Pre-assigned Orders and Sequential Organization. Our study revealed that participants frequently prepared objects in a **specific order based on their sequential requirements** (P2, P5) [**Position**]. This approach involved completing preparation tasks before embarking on the actual cooking process, thus optimizing their time and effort. P2 exemplified this by arranging the bowl of oil in front of the onions, followed by the meat, to denote the order of preparing the dish: “*Pre-arranging all the containers would save me time and effort during execution by minimizing the need to recheck the objects in the bowl.*” Furthermore, beyond arranging ingredients in the order they would be used, participants also **organized seasonings according to their frequency** of use (P9, P11) [**Similarity and Duplicates**]. P9 elaborated:

“I typically position the seasonings I use daily closer to the gas range, while placing others in cabinets or higher shelves.”

Additionally, participants maintained the **spatial arrangement of objects** even during the post-cooking cleaning process (P4, P5, P7, P12) [**Proximity and Grouping**]. This practice served a dual purpose, as explained by P5:

“The cleaning process not only ensures cleanliness but also guarantees that all objects are returned to their designated positions, preventing difficulties in locating ingredients during subsequent cooking sessions.”

5.3.2 Step and Purpose Combination. In pursuit of greater cooking efficiency, our participants skillfully combined specific cooking steps and harnessed objects for multiple purposes, simplifying their culinary endeavors, which reduced the need to obtain contextual information of multiple positions as well as differentiating different objects [**Position**][**Similarity and Duplicates**]. These strategies included merging cooking steps and employing objects with dual functionalities. Participants deliberately **integrated particular cooking steps and repurposed objects, effectively streamlining their culinary processes**. For instance, P3 favored using a pot as a multi-purpose container, consolidating ingredients in the wok before cooking. P3 elucidated: “*I adopt this approach to reduce the complexity and hassle associated with employing multiple containers.*” Similarly, P4 opted to place pepper and other seasonings inside the wok with oil before igniting the flame, effectively condensing multiple steps into a single action, which was typically divided in conventional recipes. Furthermore, our participants exhibited resourcefulness by **utilizing objects with dual functionalities to diminish the need for additional tools**. For instance, P11 employed a large bowl both as a wok lid and as a receptacle for washing vegetables. P11 explained the practicality of this approach: “*Utilizing the large bowl as a wok lid not only eliminates the necessity for a separate lid but also serves as a convenient vessel for washing various vegetables.*”

5.3.3 Reducing the Need for Precise Movement. In their quest for enhanced convenience and reduced demand for precision and accuracy, our participants expressed a preference for **minimizing complex 3D spatial actions** [**Orientation**]. They identified strategies that streamlined their actions, such as adjusting their knife-handling technique and adopting a pragmatic approach to discarding waste. Several participants indicated a desire for simplified knife handling, opting for a 2D movement approach by gripping the knife’s back edge rather than the blade’s tip (P4). This technique allowed for more straightforward and manageable motion when manipulating the knife during culinary tasks.

To further simplify their kitchen activities, our participants highlighted the practice of initially disposing of waste items in the sink (P6) [**Orientation**]. This approach was especially advantageous for items that required precise disposal, as participants found it challenging to accurately target a conventional waste bin placed at ground level. By contrast, the sink offered a more accessible, waist-level receptacle, reducing the need for precision and diminishing concerns about transferring waste to a traditional trash bin after cooking. P6 explained: “*This would reduce my effort of accurately throwing the garbage inside the bin...The water sink is big enough and at my waist level so I can easily throw things in it without much effort and worry about transferring them into the trash bin later after cooking.*”

6 FINDINGS: CONTEXTUAL INFORMATION COMMUNICATION AND DEPLOYMENT CONSIDERATIONS (RQ3)

In this section, we show our findings about contextual information presentation and deployment considerations by people with vision impairments during cooking activities. We present participants' preferences for information granularity in communication (Section 6.1), communication modality (Section 6.2), and form factor considerations for future technologies (Section 6.3).

6.1 Preferred Information Granularity and Level of Detail

6.1.1 Providing precise spatial descriptions. Participants preferred specific, stationary spatial references to describe object positions, using objects in the kitchen as landmarks. P4 elaborated:

"I typically say that the salt is on the left-hand side of the gas range, or the oyster sauce is on the windowsill. I prefer not to know if my oyster sauce is close to my sugar, because other people might place it in different spots."

Moreover, participants requested detailed spatial references. P1, for instance, suggested describing an object as being on the third shelf of the fridge door, rather than simply mentioning that it is on a shelf. Best practices here depend on layout: a kitchen with only window has objects "on the windowsill", while more details are needed for a kitchen with multiple windows. In some cases, participants combined multiple objects to describe a location, such as "the bowl is inside the cabinet at the left-hand side of my gas range." (P1).

6.1.2 Limiting verbosity. Participants expressed a dislike for systems that chatter constantly, as it could distract them from tasks that require focused auditory attention. P2 conveyed this sentiment:

"I don't want the system to constantly provide verbal updates for every new object it encounters. It should provide contextual information thoughtfully."

Participants requested that notifications be provided when most needed, and taking into account the user's current location (P4, P7, P10). P7 elaborated:

"I don't require precise navigation to a specific reference point in my kitchen, such as the position of my sink. I can manage that on my own. What I need is a brief notification to confirm that I've reached the desired position."

6.2 Preferred Communication Modalities

6.2.1 Combining Speech and Non-Speech Audio Feedback. During the post-interviews, we asked participants about their preferred modes of receiving information in the kitchen. A majority of participants preferred non-speech audio notifications, such as a familiar "beep-beep-beep" sound, as these would be less distracting. This preference was primarily motivated by the desire to minimize cognitive load during cooking, as noted in Section 5.1.2. P4, P8, and P10 mentioned the advantage of less distracting audio feedback,

particularly when engaged in conversation or consuming other auditory content. P10 pointed out:

"I prefer non-verbal sounds over human voice notifications as they are less disruptive while I'm cooking. If I'm in the middle of cooking and listening to a lecture, having a human voice as a notification can disrupt my experience."

Participants noted that speech was still needed to convey details, but they requested control over when speech feedback was offered. As articulated by P11:

"I wouldn't want the system to speak to me in a human voice unless I specifically request it. Typically, not during active cooking, but perhaps I might ask for more detailed information before or after I perform certain actions, such as inquiring about the quality of the dishes I've prepared."

Overall, participants hoped for a balance between less distracting audio cues and on-demand speech feedback.

6.2.2 Offering Proactive Reminders. Furthermore, our participants desired proactive reminders around key topics such as safety- and health-related information (P8, P11), as noted in Sections 4.2.1 and 4.2.2. Participants stressed the importance of being able to edit and customize reminders. Reminders could be triggered by events in the kitchen or based on the user's location. P12 elaborated on this concept, likening it to an indoor navigation system:

"This would resemble an indoor navigation system that can automatically provide me with the references I've designated as relevant to my current location."

6.2.3 Increasing Environmental Awareness. When discussing the benefits of using a wearable camera to track the user's activities in the kitchen, nine participants emphasized that such a system would be tolerable, but that they also might need awareness of events happening outside the view of a wearable camera. P5 articulated this need:

"I'm unsure if I can always have the tracking system facing the exact objects or reference points. Therefore, I want the system to be capable of providing me with contextual information even when I'm not directly facing it."

6.3 Deploying New Technologies in the Kitchen

6.3.1 Kitchen Deployment Considerations. During the contextual inquiry, we inquired about our participants' interest and concerns related to technology that could track contextual information. Our participants raised concerns related to introducing new technology into cooking, including the risk of water splatter, exposure to fire, oil-proofing, battery life, and dirt resistance. P9 also expressed concerns about the potential high cost of such a system, suggesting that additional support could be provided in existing mobile device hardware.

6.3.2 Device Form Factors and Location. Assuming new technology became available to support activities in the kitchen, we asked participants whether they would prefer a stationary system installed in the kitchen or a body-worn system. Out of the 12 participants, 11

expressed a preference for body-worn systems due to factors such as ease of installation, extensibility, and cost efficiency. P4 raised questions about the challenges of installing technology themselves. P7 mentioned that new technology might also be used for other purposes, such as indoor navigation. One participant favored a stationary camera due to privacy concerns associated with wearing such a system during other activities (P12).

In addition to this feedback, we conducted an evaluation in which participants wore the camera in various body positions, including the chest, head, and wrist, and asked about their preferences for device location. Eight participants indicated a preference for a chest-worn camera, while four favored a head-worn camera. Those in favor of the chest-worn position cited its non-obtrusive nature, expressing concern that a head- or hand-worn device might collide with obstacles (P5). Participants who preferred the head-worn position emphasized the advantages of greater freedom of movement when engaged in cooking tasks.

7 DISCUSSION AND FUTURE DIRECTIONS

In this section, we explore how the contextual needs identified in this study might be integrated into future technology such as AI-powered kitchen assistants.

7.1 Importance and Opportunities for AI to Improve Contextual Awareness

Prior research has explored various ways of adopting AI-powered systems to identify objects, such as using computer vision with overhead RGB-D cameras [56] or recognizing objects in mobile device images [2, 47]. While these systems typically detect objects, our study notes the importance of both identifying objects and describing the contextual information of an object. For example, existing systems mostly identify the object name (e.g., milk bottle) or a group name of objects (e.g., vegetables), instead of providing more contextual information to people (e.g., distance to the object, expiration date of food). Based on our findings on what and how contextual information should be presented, there is also an opportunity for these AI-powered systems to provide more precise spatial descriptions that are customized to the user's workspace. Beyond kitchen contexts specifically, we recommend future research to also explore other scenarios that embedding specific contextual information of objects can support people with vision impairments with agency and autonomy (e.g., art museum [41], grocery store [65], makeup [40]).

7.2 Creating Smart Objects in the Kitchen

In our study, we found that visually impaired participants often substituted touch for visual information, and did additional work to make recognition by touch easier, such as using different bottle shapes for different ingredients (Section 5.2). If users are already choosing or augmenting the shapes of objects, it might be feasible to expect users to attach tags to objects that could improve recognition [23, 39]. Using more complex tagging methods such as 3D printed models [54] or sensor-enabled tags [57] could provide additional contextual information. This leads to future fabrication research to consider exploring: 1) tagging methods that are sustainable and deformable so that they can be attached to different kitchen objects,

2) platforms that support people with vision impairments to create customized 3D objects for their home, 3) low-cost embedded systems to track and report back about item status.

7.3 Augmenting the Kitchen or the User

Beyond customizing contextual information of objects, we also uncovered the importance of managing the space (Section 4.1.3) and knowing the internal state of objects (Section 4.1.5), which often require them to leverage touch to scan through the space (Section 5.1.1). Prior research has explored approaches to **augment the kitchen space**, such as by instrumenting a kitchen with cameras, microphones, and motion tracking [49, 64]. These approaches should also work for people with vision impairments, and may provide even more benefit as they may help address accessibility challenges in the kitchen. Creating and deploying such systems would need to acknowledge that visually impaired users may move and act differently within the kitchen space, and may have particular concerns around issues such as spilling ingredients or tracking cooking status.

Along with augmenting the kitchen space, most of our participants were open to the idea of using **wearable devices** (e.g., body-worn cameras or smartwatches) on themselves to track their activities and provide contextually-relevant suggestions (Section 6.3). For example, we found that tracking the internal state of the space and objects can be highly visual and it brings opportunities to create systems to visually check the internal state of objects through user definitions (e.g., dumplings floating indicate doneness). To understand visual information, devices with worn cameras could leverage pre-trained Vision-Language Models (VLMs) [10, 15, 19, 66] to track the status of objects and answer user questions (VQA), such as safety-related questions as well as relative positions and state of objects (e.g., material, quantity) [19]. Such a system could also provide proactive notifications, such as noting if an object has moved or if the space has been rearranged by another user of that kitchen.

7.4 Tracking Activities in the Background

Participants noted that they would be interested in knowing the status of the kitchen even when they had moved outside of that space (Section 6.2.3). Implicit tracking systems [21, 25, 29] could discreetly and continuously monitor contextual information, such as the location of objects, while the user is engaged in various activities within the kitchen [37]. This approach would alleviate the need for users to consciously track objects in the kitchen at all times. Future research should consider 1) achieving comprehensive coverage within diverse kitchen spaces, 2) the practicality of addressing challenges related to power consumption and storage capacity [37], 3) the interactive interface for people with vision impairments to pre-assign object of interests.

7.5 Multimodal Interaction in the Kitchen

Participants' activities in the kitchen often leveraged multiple sensory modes at once and in concert (e.g., touch, sound, and smell) (Section 5.1). Participants also experienced overload and related challenges during these tasks (Section 5.1). This type of multimodal interaction is known in HCI to support users with diverse abilities

[7, 42, 45, 50, 51, 53]. As users with vision impairments already interact multimodally in the kitchen, technology that supports these users should also be multimodal and adapted to their existing ways of performing tasks (Section 5.1). We noted that participants' abilities to engage the environment were sometimes affected by context, for example, using touch is not always feasible when the user's hands are dirty or greasy, which suggests that future systems should be context-aware and should adapt to the user's activities and current state.

8 LIMITATIONS

In our research, our participant pool consisted exclusively of individuals with prior cooking experience. Users with no cooking experience would also benefit from this work, but would likely encounter different problems. All of our participants were either legally or completely blind and did not use any type of vision for assistance while cooking. However, we believe some people with low vision might have different practices and challenges of obtaining objects' contextual information during cooking processes and we suggest future research could further conduct studies with a broader range of visually impaired people on this aspect. Moreover, it is worth noting that our study participants were recruited locally in mainland China. Consequently, individuals from different geographic regions or diverse cultural backgrounds may exhibit varying behaviors and unique contextual information needs within the kitchen setting [38]. Extending this work to additional geographic and cultural contexts would likely lead to additional, and complementary, insights.

9 CONCLUSION

While emerging technologies could support visually impaired users in activities of daily living, it is essential that these technologies are designed with a robust understanding of how users actually perform tasks. Our contextual inquiry study identified the primary sources of contextual information used by visually impaired cooks, and revealed challenges encountered by the visually impaired users when cooking. Our study also explored preferences and concerns related to the introduction of new technology into the kitchen, and identified preferences for how such technologies should communicate with users (with key issues including mode of communication and levels of verbosity). Finally, we propose future research directions to increase contextual awareness for individuals with vision impairments when working in the kitchen. Our work provides initial steps towards enabling individuals with vision impairments to access and track contextual information concerning objects during culinary endeavors.

ACKNOWLEDGMENTS

This work was funded in part by Google Research and NSERC Fellowship PGS-D-557565-2021. We especially thank all participants who accommodated the contextual inquiry in their kitchens. We would also like to sincerely thank Yunfeng Shi and Ziyue Hu for participant recruitment and study support.

REFERENCES

- [1] Karissa Bell. [n.d.]. Microsoft's 'talking camera' app for the blind isn't as magical as it sounds | Mashable. <https://mashable.com/article/microsoft-seeing-ai-app-for-blind>. (Accessed on 09/10/2023).
- [2] BeMyEyes. 2023. Introducing: Be My AI. <https://www.bemyeyes.com/blog/introducing-be-my-ai>. (Accessed on 11/22/2023).
- [3] Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. *interactions* 6, 1 (1999), 32–42.
- [4] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, et al. 2010. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*. 333–342.
- [5] Marie Claire Bilyk, Jessica M Sontrop, Gwen E Chapman, Susan I Barr, and Linda Mamer. 2009. Food experiences and eating patterns of visually impaired and blind people. *Canadian Journal of Dietetic practice and research* 70, 1 (2009), 13–18.
- [6] Rubén Blasco, Álvaro Marco, Roberto Casas, Diego Cirujano, and Richard Picking. 2014. A smart kitchen for ambient assisted living. *Sensors* 14, 1 (2014), 1629–1653.
- [7] Jullien Bouchet and Laurence Nigay. 2004. ICARE: a component-based approach for the design and development of multimodal interfaces. In *CHI'04 extended abstracts on Human factors in computing systems*. 1325–1328.
- [8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [9] Minsuk Chang, Léonore V Guillaín, Hyeunghshik Jung, Vivian M Hare, Juho Kim, and Maneesh Agrawala. 2018. Recipescape: An interactive tool for analyzing cooking instructions at scale. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
- [10] Yu-Wei Chao, Yunfan Liu, Xieyang Liu, Huayi Zeng, and Jia Deng. 2018. Learning to detect human-object interactions. In *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 381–389.
- [11] Jen-Hao Chen, Peggy Pei-Yu Chi, Hao-Hua Chu, Cheryl Chia-Hui Chen, and Polly Huang. 2010. A smart kitchen for nutrition-aware cooking. *IEEE Pervasive Computing* 9, 4 (2010), 58–65.
- [12] Antonella De Angeli, Uday Athavankar, Anirudha Joshi, Lynne Coventry, and Graham I Johnson. 2004. Introducing ATMs in India: a contextual inquiry. *Interacting with Computers* 16, 1 (2004), 29–44.
- [13] Bryan Dosono, Jordan Hayes, and Yang Wang. 2015. "I'm Stuck!": A Contextual Inquiry of People with Visual Impairments in Authentication. In *Eleventh Symposium On Usable Privacy and Security (SOUPS) 2015*. 151–168.
- [14] Bengisu Dost. [n.d.]. How can we improve Microsoft's "Seeing AI" application? | by Bengisu Dost | Medium. <https://bengisudost.medium.com/how-can-we-improve-microsofts-seeing-ai-application-4cb4b074e0d6>. (Accessed on 09/10/2023).
- [15] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. 2023. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378* (2023).
- [16] Thomas Erickson and David W McDonald. 2007. *HCI Remixed: Reflections on Works That Have Influenced the HCI Community*. MIT Press.
- [17] Maurizio Ficoelli and Goldie Nejat. 2012. The design of an interactive assistive kitchen system. *Assistive Technology* 24, 4 (2012), 246–258.
- [18] Giovanni Fusco, Ender Tekin, Richard E Ladner, and James M Coughlan. 2014. Using computer vision to access appliance displays. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. 281–282.
- [19] Jensen Gao, Bidipta Sarkar, Fei Xia, Ted Xiao, Jiajun Wu, Brian Ichter, Anirudha Majumdar, and Dorsa Sadigh. 2023. Physically Grounded Vision-Language Models for Robotic Manipulation. *arXiv preprint arXiv:2309.02561* (2023).
- [20] GoPro. 2023. GoPro HERO11 Black Action Camera (Waterproof + Stabilization). <https://gopro.com/en/us/shop/cameras/hero11-black/CHDXH-111-master.html>. (Accessed on 09/04/2023).
- [21] Rúben Gouveia and Evangelos Karapanos. 2013. Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2921–2930.
- [22] Anhong Guo, Xiang'Anthony' Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P Bigham. 2016. Vizlens: A robust and interactive screen reader for interfaces in the real world. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 651–664.
- [23] Anhong Guo, Jeeun Kim, Xiang'Anthony' Chen, Tom Yeh, Scott E Hudson, Jennifer Mankoff, and Jeffrey P Bigham. 2017. Facade: Auto-generating tactile interfaces to appliances. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5826–5838.
- [24] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. 2018. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3608–3617.
- [25] Steve Hodges, Emma Berry, and Ken Wood. 2011. SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory* 19, 7 (2011),

- 685–696.
- [26] Karen Holtzblatt and Hugh Beyer. 1995. Contextual Design: using customer work models to drive system design. In *Conference companion on human factors in computing systems*. 391–392.
 - [27] Nabila Jones, Hannah Elizabeth Bartlett, and Richard Cooke. 2019. An analysis of the impact of visual impairment on activities of daily living and vision-related quality of life in a visually impaired adult population. *British Journal of Visual Impairment* 37, 1 (2019), 50–63.
 - [28] Holtzblatt Karen and Jones Sandra. 2017. Contextual inquiry: A participatory technique for system design. In *Participatory design*. CRC Press, 177–210.
 - [29] Shuaib Karim, Amin Andjomshoa, and A Min Tjoa. 2006. Exploiting sensecam for helping the blind in business negotiations. In *Computers Helping People with Special Needs: 10th International Conference, ICCHP 2006, Linz, Austria, July 11-13, 2006. Proceedings 10*. Springer, 1147–1154.
 - [30] Mohammad Kianpisheh, Franklin Mingzhe Li, and Khai N Truong. 2019. Face recognition assistant for people with visual impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
 - [31] Yonghee Kim, Youngju Park, and Jeongil Choi. 2017. A study on the adoption of IoT smart home service: using Value-based Adoption Model. *Total Quality Management & Business Excellence* 28, 9-10 (2017), 1149–1165.
 - [32] Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner. 2012. *Video analysis: Methodology and methods*. Peter Lang Frankfurt am Main.
 - [33] A König and K Thongpull. 2015. Lab-on-Spoon—a 3-D integrated hand-held multi-sensor system for low-cost food quality, safety, and processing monitoring in assisted-living systems. *Journal of Sensors and Sensor Systems* 4, 1 (2015), 63–75.
 - [34] Eliza Kostyra, Sylwia Żakowska-Biemans, Katarzyna Śniegocka, and Anna Piotrowska. 2017. Food shopping, sensory determinants of food choice and meal preparation by visually impaired people. Obstacles and expectations in daily food experiences. *Appetite* 113 (2017), 14–22.
 - [35] Kazuma Kusu, Nozomi Makino, Takamitsu Shioi, and Kenji Hatano. 2017. Calculating Cooking Recipe's Difficulty based on Cooking Activities. In *Proceedings of the 9th Workshop on Multimedia for Cooking and Eating Activities in conjunction with The 2017 International Joint Conference on Artificial Intelligence*. 19–24.
 - [36] Jinna Lei, Xiaofeng Ren, and Dieter Fox. 2012. Fine-grained kitchen activity recognition using rgb-d. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 208–211.
 - [37] Franklin Mingzhe Li, Di Laura Chen, Mingming Fan, and Khai N Truong. 2019. FMT: A Wearable Camera-Based Object Tracking Memory Aid for Older Adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–25.
 - [38] Franklin Mingzhe Li, Di Laura Chen, Mingming Fan, and Khai N Truong. 2021. "I Choose Assistive Devices That Save My Face" A Study on Perceptions of Accessibility and Assistive Technology Use Conducted in China. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
 - [39] Franklin Mingzhe Li, Jamie Dorst, Peter Cederberg, and Patrick Carrington. 2021. Non-visual cooking: exploring practices and challenges of meal preparation by people with visual impairments. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–11.
 - [40] Franklin Mingzhe Li, Franchesca Spektor, Meng Xia, Mina Huh, Peter Cederberg, Yuqi Gong, Kristen Shinohara, and Patrick Carrington. 2022. "It Feels Like Taking a Gamble": Exploring Perceptions, Practices, and Challenges of Using Makeup and Cosmetics for People with Visual Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.
 - [41] Franklin Mingzhe Li, Lotus Zhang, Maryam Bandukda, Abigale Stangl, Kristen Shinohara, Leah Findlater, and Patrick Carrington. 2023. Understanding Visual Arts Experiences of Blind People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
 - [42] Michael Xieyang Liu. 2023. *Tool Support for Knowledge Foraging, Structuring, and Transfer During Online Sensemaking*. Ph.D. Dissertation. Ph. D. Dissertation. Carnegie Mellon University. <http://reports-archive.adm>
 - [43] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A Myers. 2019. Unakite: Scaffolding developers' decision-making using the web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 67–80.
 - [44] Michael Xieyang Liu, Aniket Kittur, and Brad A Myers. 2021. To reuse or not to reuse? A framework and system for evaluating summarized knowledge. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–35.
 - [45] Michael Xieyang Liu, Andrew Kuznetsov, Yongsung Kim, Joseph Chee Chang, Aniket Kittur, and Brad A Myers. 2022. Wigglyte: Low-cost Information Collection and Triage. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
 - [46] Haley MacLeod, Cynthia L Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding blind people's experiences with computer-generated captions of social media images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5988–5999.
 - [47] Microsoft. 2023. Seeing AI App from Microsoft. <https://www.microsoft.com/en-us/ai/seeing-ai>. (Accessed on 09/03/2023).
 - [48] Tim Morris, Paul Blenkhorn, Luke Crossey, Quang Ngo, Martin Ross, David Werner, and Christina Wong. 2006. Clearspeech: A display reader for the visually handicapped. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14, 4 (2006), 492–500.
 - [49] Patrick Olivier, Guangyou Xu, Andrew Monk, and Jesse Hoey. 2009. Ambient kitchen: designing situated services using a high fidelity prototyping environment. In *Proceedings of the 2nd international conference on pervasive technologies related to assistive environments*. 1–7.
 - [50] Sharon Oviatt, Björn Schuller, Philip Cohen, Daniel Sonntag, and Gerasimos Potamianos. 2017. *The handbook of multimodal-multisensor interfaces, volume 1: Foundations, user modeling, and common modality combinations*. Morgan & Claypool.
 - [51] Jennifer Preece, Yvonne Rogers, and Helen Sharp. 2015. *Interaction Design: Beyond Human-Computer Interaction* (4 ed.). Wiley, Hoboken, NJ.
 - [52] Mary Elizabeth Raven and Alicia Flanders. 1996. Using contextual inquiry to learn about your audiences. *ACM SIGDOC Asterisk Journal of Computer Documentation* 20, 1 (1996), 1–13.
 - [53] Leah M Reeves, Jennifer Lai, James A Larson, Sharon Oviatt, TS Balaji, Stéphanie Buisine, Penny Collings, Phil Cohen, Ben Kraal, Jean-Claude Martin, et al. 2004. Guidelines for multimodal user interface design. *Commun. ACM* 47, 1 (2004), 57–59.
 - [54] Lei Shi, Yuhang Zhao, and Shiri Azenkot. 2017. Designing interactions for 3D printed models with blind people. In *Proceedings of the 19th international acm sigaccess conference on computers and accessibility*. 200–209.
 - [55] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.
 - [56] Sebastian Stein and Stephen J McKenna. 2013. Combining embedded accelerometers with computer vision for recognizing food preparation activities. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 729–738.
 - [57] Ryo Suzuki, Abigale Stangl, Mark D Gross, and Tom Yeh. 2017. Fluxmarker: Enhancing tactile graphics with dynamic tactile markers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 190–199.
 - [58] TapTapSee. 2023. TapTapSee - Blind and Visually Impaired Assistive Technology - powered by CloudSight.ai Image Recognition API. <https://taptapseeapp.com/>. (Accessed on 09/10/2023).
 - [59] Ender Tekin, James M Coughlan, and Huiying Shen. 2011. Real-time detection and reading of LED/LCD displays for visually impaired persons. In *2011 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 491–496.
 - [60] YingLi Tian, Chucai Yi, and Aries Arditi. 2010. Improving computer vision-based indoor wayfinding for blind persons with context information. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July 14-16, 2010, Proceedings, Part II 12*. Springer, 255–262.
 - [61] Maarten Van Someren, Yvonne F Barnard, and J Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: Academic Press* 11 (1994), 29–41.
 - [62] Trieu Minh Vu and Riva Khanna. 2018. Application of artificial intelligence in smart kitchen. *International Journal of Innovative Technology and Interdisciplinary Sciences* 1, 1 (2018), 1–8.
 - [63] Ru Wang, Nihan Zhou, Tam Nguyen, Sanbrita Mondal, Bilge Mutlu, and Yuhang Zhao. 2023. Practices and Barriers of Cooking Training for Blind and Low Vision People. (2023).
 - [64] Nikola Žarić, Milutin Radonjić, Nikola Pavličević, and Sanja Paunović Žarić. 2021. Design of a kitchen-monitoring and decision-making system to support AAL applications. *Sensors* 21, 13 (2021), 4449.
 - [65] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 73–84.
 - [66] Kaiyang Zhou, Jingkan Yang, Chen Change Loy, and Ziwei Liu. 2022. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16816–16825.