



HHS Public Access

Author manuscript

DIS (Des Interact Syst Conf). Author manuscript; available in PMC 2025 January 04.

Published in final edited form as:

DIS (Des Interact Syst Conf). 2023 ; 2023: 1810–1825. doi:10.1145/3563657.3596019.

Are Two Heads Better than One? Investigating Remote Sighted Assistance with Paired Volunteers

Jingyi Xie,

Pennsylvania State University, University Park, PA, USA

Rui Yu,

Pennsylvania State University, University Park, PA, USA

Kaiming Cui,

Pennsylvania State University, University Park, PA, USA

Sooyeon Lee,

New Jersey Institute of Technology, Newark, NJ, USA

John M. Carroll,

Pennsylvania State University, University Park, PA, USA

Syed Masum Billah

Pennsylvania State University, University Park, PA, USA

Abstract

Remote Sighted Assistance (RSA) is a popular smartphone-mediated aid for people with blindness, where a sighted individual converses with a blind individual in a one-on-one (1:1) session. Since sighted assistants outnumber blind individuals (13:1), this paper investigates what happens when more than one sighted individual assists a single blind individual in a session. Specifically, we propose paired-volunteer RSA, a new paradigm where two sighted volunteers assist a single user with blindness. We investigate the feasibility, desirability, and challenges of this paradigm and explore its opportunities. Our study with 8 sighted volunteers and 9 blind users reveals that the proposed paradigm extends the one-on-one RSA to cover a broader range of more intellectual and experiential tasks, providing new and distinctive opportunities in supporting complex, open-ended tasks (e.g., pursuing hobbies, appreciating arts, and seeking entertainment). These opportunities can not only enrich the blind users' quality of life and independence but also offer a fun and engaging experience for the sighted volunteers. The study also reveals the costs of extended collaboration in this paradigm. Finally, we synthesize a taxonomy of tasks where the proposed RSA paradigm can succeed and outline how HCI researchers and system designers can realize this paradigm.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

jzx5099@psu.edu .

Keywords

People with visual impairments; paired-volunteer remote sighted assistance; awareness; group work; helper-worker model

1 INTRODUCTION

Remote sighted assistance (RSA) is a popular smartphone-mediated conversational assistance for people with blindness [50]. During an RSA session, a user with blindness (“*user*” for short) connects with a remotely-located sighted human assistant (“*assistant*” for short); the user shares their live camera feed with the sighted assistant, who interprets the video feed as appropriate and converses with the user to provide assistance. RSA services, such as Be My Eyes (BME) [6] and Aira [2], have been successful for tasks ranging from low-stake, everyday inquiries (e.g., what is the color of a dress) [12, 20] to high-stake, complex navigational tasks (e.g., navigating airports) [51, 79].

Despite their versatility and varied complexity, tasks performed on RSA services share a common trait – they are all basic and objective [50], with solutions that are either known or universally agreed upon. This limits the potential of RSA, as the rich communication and collaboration between users and assistants could enable the execution of open-ended tasks (e.g., knitting [27]), which could enrich blind users’ lives and promote overall well-being.

This potential is particularly evident in unpaid, volunteer-based RSA services like Be My Eyes, which has a significant pool of sighted assistants – over 6 million worldwide – ready to assist approximately 0.45 million blind and low-vision users [6]. The ratio of registered users to assistants stands at a notable 1 to 13, suggesting that for each registered user, there are 13 registered volunteers available to assist. This imbalance suggests two things: i) *the volunteer resource is likely underutilized*, as volunteers on BME can wait from one day to several months between calls [5]; and ii) *sighted volunteers are motivated to help blind individuals*, possibly driven by philanthropy or genuine curiosity about the experiences and perspectives of blind users.

The large ratio of users to volunteers in RSA, coupled with the motivation of sighted volunteers to assist blind individuals and the need to support open-ended tasks, encourages us to explore the potential for more collaborative interactions in RSA service provision. In response, we propose a novel RSA paradigm, *paired-volunteer RSA (paired-RSA)*, where two sighted volunteers assist a single user in a session. This could potentially amplify the societal benefit by enabling direct interaction, collaboration, and mutual understanding between individuals with different visual abilities.

Compared to the traditional one-on-one RSA, which employs a single communication channel (as shown in Fig. 1), we hypothesize that the proposed paired-RSA would likely be more capable, collaborative, and engaging. It entails three-way conversations (three communication channels), which could lead to better resource utilization. However, this paradigm might also introduce unwanted side effects and may not be appropriate for

certain tasks. Consequently, this paper aims to understand the feasibility, opportunities, and challenges of the paired-RSA paradigm.

In particular, we investigate the following *research questions*:

RQ₁. *What are new, distinctive opportunities offered by collaborative paired-RSA interactions?*

RQ₂. *What are the benefits and costs of this extended collaboration?*

To answer the above research questions, we conducted an exploratory study. First, we ideated a list of candidate tasks that could go beyond the frequently-requested tasks in the current RSA. We then consulted 2 domain experts to refine our task list (Section 3). These tasks were sent to 9 recruited users before the experimental sessions, who were allowed to choose any tasks that they want to get help with. Finally, we conducted 11 online paired-RSA sessions of various tasks (Section 4).

Our findings (Section 5) reveal that paired-RSA opens up new collaborative possibilities for volunteers injecting more perspectives and more knowledge, thus extending RSA into a broader coverage of more intellectual and experiential tasks. Compared with one-on-one RSA, paired-RSA is more beneficial in entertainment tasks (e.g., developing hobbies, appreciating artwork, and crafting); tasks requiring multiple opinions or perspectives (e.g., matching outfits, applying makeup); and tasks with high cognitive load only if parallelism applied (e.g., navigating in traffic scenes).

In this small-group collaboration, the audio-video hybrid-channel communication between three partners entails social engagements, and establishes and maintains common grounds about what their counterparts are doing. As assistance providers, volunteers create impromptu roles through turn-taking and division of labor to cultivate and reinforce their synergy. However, task demands, partners' personalities, and expertise might stray from their plans or expectations. This divergence could undermine collaboration and culminate in a breakdown, evident by the dispensable input from the second volunteer in simple objective tasks (e.g., reading mail) and the overload of the audio channel in time-sensitive, high-risk navigational tasks (e.g., grocery shopping).

We synthesized a taxonomy of tasks where paired-RSA can succeed (Section 6), suggested categories of extensions to RSA, and explored one token for every type. By investigating more examples, HCI researchers and system designers could realize this paradigm and utilize a set of scenarios where paired-RSA is feasible and desirable. Furthermore, we examined the usefulness of paired-RSA through indicators of success in remote collaboration [59] and discussed factors that could affect its performance. It lays the foundation for multi-way, sighted-blind cooperation and calls for further research into smooth transitions from RSA to paired-RSA.

2 BACKGROUND AND RELATED WORK

Living independently remains a significant challenge for individuals with blindness or low vision [64]. To manage, they utilize a range of aids, including their enhanced sensory

capabilities [43, 77], orientation and mobility (O&M) skills [78], traditional mobility tools (e.g., white canes and guide dogs), and tech-based assistance. Such technology encompasses smartphone-based apps that leverage sensors, mapping services, computer vision, and artificial intelligence (e.g., [3, 9, 16, 35, 42, 52, 57, 66, 68, 72, 81]).

Despite their sophistication, many of these technologies are not universally viable due to their lab-based testing, cost, unreliability, complexity, or distribution issues [28, 78]. Therefore, in-person sighted assistance often proves the most effective aid for those with visual impairments. However, dependence on sighted individuals—usually friends or family—can be burdensome and impractical [36]. As a result, blind and low-vision individuals increasingly rely on remote sighted assistants through video chat-like calls, who offer readily available help without requiring a prior connection, presenting a simple solution to this ongoing challenge.

The subsequent sections provide an overview of the current status of remote sighted assistance (RSA) for individuals with blindness or low vision. Furthermore, we situate our research within the extensive body of literature on remote collaborative work.

2.1 Remote Sighted Assistance (RSA) Services

Evolution of RSA technology and communication channels.—The proliferation of smartphones and high-speed mobile Internet enabled RSA services to become an increasingly important means of assistance for the blind population. RSA services continuously adopted improved technology and communication channels. For instance, early prototypes used images [15, 46], audio [61], one-way video from webcams [19] or portable digital cameras [19, 31], whereas the recent ones use two-way, video-chat-like interface with smartphones or smart glasses [2, 6, 13, 39]. Similarly, the communication channels have evolved to include texts [47], synthetic speech [61], vibrotactile feedback [24, 69], and more recently, natural conversation [2, 6, 13]. Our proposed paired-RSA benefits from the continuous improvements in technology, and, instead of a single conversational channel, it uses three channels.

Existing types of RSA services.—Currently, RSA services are facilitated by two categories of assistants: unpaid volunteers, as seen in Be My Eyes [6], and paid professional agents, like those in Aira [2]. The volunteer-based services offer a vast resource pool of sighted assistants, accessible globally and in various languages [6]. However, due to their non-profit nature, the service quality isn't guaranteed [12, 23, 25]. For instance, volunteers often have minimal training in O&M skills [20, 53, 62]. On the other hand, paid RSA services employ well-trained agents who are proficient in communication etiquette [50, 58] and high-stakes navigational tasks [21, 50, 51, 79]. However, these services are somewhat restricted in terms of the number of available languages and agents, and their professional obligation limits them from offering subjective opinions on a topic [50].

Limitations and potential advancements.—Despite differences in availability and affordability, all existing RSA services share a common limitation: they permit only one sighted volunteer or agent per call. Prior work has underscored the challenges faced by these sighted individuals, noting that remote assistance can be mentally taxing. This is due

to a lack of familiarity with the blind user's current physical environment, the scarcity of detailed indoor maps, the difficulty of continuously tracking and orienting users within their surroundings, the challenge of estimating object depth, detecting landmarks and obstacles in the blind user's camera feed, and the need to interpret and deliver visual information in real-time through conversations [44, 51]. Previous research has proposed the use of computer vision (CV) technologies, such as 3D map construction, object annotations, real-time localization, and video stream augmentation, to address these challenges [21, 79]. Preliminary results from the lab prototype [80] indicate that CV-mediated RSA could enhance the sighted assistants' ability to anticipate user needs and reduce their mental workload.

Focus of this study.—This study deviates from previous research by examining the feasibility, potential benefits, and challenges associated with incorporating an extra sighted volunteer into the existing RSA structure. We anticipate this inquiry will provide valuable insights into the benefits, such as receiving subjective perspectives while minimizing biases, potential challenges, including conflicts among sighted volunteers, and identification of task types that fit this paradigm.

2.2 RSA: A Helper-Worker Collaboration Model

The “helper-worker” model is a well-established framework for understanding distinct types of cooperative work [29, 30, 32, 45, 63]. In this model, collaboration typically involves one participant — the “helper” — providing support to another participant — the “worker” — to help them accomplish a task. This model is particularly relevant when the worker lacks certain specialized knowledge, skills, or abilities required to perform the task independently. Instead of directly executing the task, the helper offers guidance, advice, and clarification, enabling the worker to effectively complete the task. For instance, a specific application of this model is the RSA interaction, where the worker is visually impaired and the helper aids them in performing visual tasks with the help of technology [21, 49–51, 79, 80].

The significance of effective communication.—The success of the “helper-worker” model is strongly contingent on effective communication between the two participants. The helper must fully comprehend the worker's abilities, limitations, and the specifics of the task to provide suitable assistance. Conversely, the worker must be able to effectively convey their needs, progress, and any challenges they encounter during the task execution. This mutual exchange of information is pivotal in establishing and maintaining common ground [26, 33, 34], thereby facilitating efficient collaboration and joint activities.

Individuals use natural conversations, non-verbal cues (e.g., head nods, facial expressions), and shared visual context (e.g., camera viewport, shared screen) for monitoring task status, other's actions, the joint focus of attention, and partner comprehension [22]. If they collaborate remotely, video systems are more effective than audio-only systems [45] despite restricting available cues compared to face-to-face interaction [76]. For instance, a broad range of camera views or controls can distract the helper and make establishing a joint focus difficult. Conversely, a narrow field of view might hinder effective monitoring [63].

Expanding the helper-worker model.—In this paper, we explore an expansion of the “helper-worker” model, where more than one helper supports a single worker. This model differs from the current RSA model, where the helper and worker only converse to establish common ground since the visual context created by the workers’ camera always streams the workspace and is not accessible to the worker. In the proposed model, the visual context can be shared among sighted helpers (e.g., shared screen, workers’ cameras), which will likely facilitate establishing common ground. Additionally, this paper delves into the dynamics of the proposed collaboration, examining how it unfolds in managing group dynamics and dealing with different task types.

2.3 Generic Frameworks of Collaborative Work

The helper-worker model is a special case of generic conceptual frameworks for the study of groups [55]. We briefly describe two key elements, group tasks and group interaction process, of this framework to position our work.

Group tasks.—McGrath [55] integrated prior task models (e.g., [38, 48, 70, 73]) and classified group tasks into four categories, forming a “task circumplex”. This classification includes: i) tasks that involve generating ideas or plans (e.g., creativity tasks, subjective tasks); ii) tasks that involve choosing a solution or plan from a set of alternatives where the correct or agreed-upon answer exists (e.g., objective tasks); iii) tasks that involve negotiating to reach a consensus (e.g., resolving conflict of viewpoints, interests, and motives); and iv) tasks that involve executing a plan or performance (e.g., competing for victory, performing for excellence). Each group task could be placed within this circumplex, providing a structured way to understand and analyze the nature of the task a group is performing. We note that most RSA tasks currently fall under the second type (objective tasks).

Group interaction process.—Group interaction is not just about the task at hand but also about the relationships among group members. Drawing on Bales’s Interaction Process Analysis [17], group interaction processes can be divided into two main categories: i) *task processes*, which are the activities directly related to the task, such as *actively* giving suggestions, opinions, and orientation; or *passively* seeking these inputs; and ii) *socio-emotional processes*, which are the activities related to building and maintaining relationships within the group. These may include *positive* expressions that demonstrate solidarity, release tension, and convey agreement or *negative* expressions that reveal antagonism, create tension, and indicate disagreement. We utilize these categories in our discussion to contextualize our findings within the literature.

3 IDEATION OF PAIRED-VOLUNTEER RSA

Drawing on prior work on RSA [12, 44, 50, 51, 79, 80], we started by gathering public data about one of the largest RSA service providers, Be My Eyes (BME), to understand the current RSA ecosystem. Our data revealed that sighted volunteers outnumber blind users by 13:1 in BME [6]. Due to this skewed distribution of volunteers to users, most volunteers remain idle and can wait several months to get a call [5]. When users initiate a call, BME adopts the “first come, first serve” policy to connect them with volunteers who are in the

same timezone, can speak in the same language, and are the first to answer the call [7]. Randomly connecting a user with a volunteer maintains the anonymity of both individuals; it simplifies the call establishment process, reducing the latency to under 15 seconds [5, 7]. Once a call is established, the user typically asks for low-stake, routine tasks [1, 8, 11, 12, 80], which we grouped into nine categories: (i) reading printed text; (ii) describing visual media, (iii) finding lost or dropped objects, (iv) kitchen assistance; (v) learning or setting up appliances; (vi) shopping; (vii) navigational assistance; (viii) assistance in pursuing hobbies and leisure activities; and (ix) style and beauty assistance. Some RSA providers (e.g., BME) allow users to call for expert services (e.g., tech support) by redirecting the call to external businesses, such as the customer service of Google Disability Support [4, 10]. Table 1 presents our task categories and a sampler of tasks frequently rendered under each category.

Our data suggest that the volunteering resource in RSA services is likely *underutilized*. Therefore, we ideate how to better utilize this resource by broadening the coverage of tasks as follows:

- **Idea 1: Pair a user with two volunteers in a call.** Instead of connecting a user with a *single* volunteer, we envision connecting a user with a *pair* of volunteers in a call to leverage the abundance of volunteers. We refer to this setting as *paired-volunteer RSA* or, in short, *paired-RSA*.
- **Idea 2: Support call scheduling.** Instead of randomly connecting a user with volunteers, we propose pre-planning tasks that (i) require volunteers' topical expertise (e.g., building computers) or (ii) take a long time to complete (e.g., describing the preseason football game). Therefore, by scheduling a call in advance, a user can give volunteers time to plan and research. In addition, it can seamlessly extend the specialized service on the RSA platforms from business domains to general task domains with volunteers who have special skills.

To refine our ideas, we consulted two domain experts: one is a high-level official involved in BME from the beginning and is familiar with all activities on this platform; another manages a large non-profit that trains BME volunteers to better communicate with and assist blind individuals during visual interpretation tasks. The latter was also a trained professional who worked as a paid RSA agent for three years prior to his current appointment. The experts encouraged our ideas and helped us contextualize them with a list of meaningful tasks.

Our envisioned tasks, presented in the third column of Table 1, are complex, real-life, and require objectivity and subjectivity. These tasks can expand the repertoire of current tasks rendered in the RSA platforms. Some of our tasks include assisting users in pursuing their hobbies and leisure activities, seeking multiple opinions, troubleshooting, and doing high-risk wayfinding. Next, we present our study.

4 STUDY: INVESTIGATING PAIRED-VOLUNTEER RSA

We conducted an exploratory study (IRB-approved) with 9 blind participants and 8 sighted volunteers to investigate the feasibility, desirability, and challenges of paired-RSA as well as to understand our research questions.

4.1 Participants

Recruiting blind users.—We recruited a total of 9 blind participants (male: 4, female: 5) through the institutional Disability Office and from our prior contacts. All blind participants were familiar with free (e.g., BME) and paid (e.g., Aira) RSA services but had no paired-RSA experience prior to this study. Their most common age group was 35–40. Three of them were students, two were unemployed, and the rest were full-time employees. Table 2 presents their demographics. Each user received a \$45 gift card per session for their time and effort.

Recruiting sighted volunteers.—We also recruited a total of 8 sighted volunteers (male: 2, female: 6), with the most common age group of 25–30. Table 3 presents their demographics. Most of them were students and had received fewer than 15 calls. V1 and V8 were notable exceptions: they had received over 50 and 300 calls, respectively. None had paired-RSA experience prior to this study. We consulted with V1 during ideation (he was the second expert in Section 3). V1 helped us send out our recruitment materials to his contacts. We also recruited some volunteers from LinkedIn. Each volunteer received a \$45 gift card per session.

4.2 Apparatus

Since current RSA services are based on smartphone apps that cannot support connecting a single user with more than one volunteer, we used the Zoom teleconferencing app to approximate paired-RSA. Zoom is similar to the RSA apps in terms of screenshot function, two-way audio connection, and video transmission functions. However, Zoom supports calling from desktop computers, screen sharing, and multi-party audio-video communication. In each session, a blind user and two volunteers joined a Zoom room with their devices (e.g., smartphones, tablets, or laptops). In addition, two to three researchers also joined in each session to observe their collaboration.

4.3 Task Design

First, we sent out the list of envisioned tasks (Table 1) to blind participants via email and asked them to choose any tasks from this list that they were interested in getting help with. They were allowed to name their own tasks for which they received limited support with the current RSA. Participants responded with one or more tasks of their choice and their availability. Based on their responses, we identified what skills volunteers should have to carry out these tasks. For example, for users who choose high-stake tasks (e.g., in-store shopping and cooking assistance), we ensured that they were connected with expert volunteers (e.g., V1 or V8) for safety. Similarly, for users who choose to read multilingual text, we matched them with volunteers with the required skill. We then reached out to our sighted volunteers, sending them the task descriptions and asking them whether they were available and interested in assisting with the tasks. In total, we administered 11 sessions. Table 4 presents the participants and tasks involved in each session. Note that all participants (blind and sighted) were aware of the tasks before the session. To increase ecological validity and generalize findings to real RSA situations, we invited participants who were strangers to each other in a session (with a few exceptions, such as sessions 1, 7, and 11).

4.4 Procedure

All sessions were conducted over Zoom and recorded after the consent. We encouraged all participants to turn on their cameras. Each session lasted for 90 minutes and was divided into three parts.

Part 1: task rendering.—The pair of volunteers were allowed to discuss their coordination plan (e.g., turn-taking) before assisting the blind participant with the requested tasks. Researchers did not interrupt this process unless they answered participants' questions. Based on the number of tasks requested and their difficulties, the duration of the first part varied from 40 to 70 minutes.

Part 2: one-on-one interviews.—After finishing the tasks, we conducted a one-on-one interview with individual participants (users and volunteers), in order to collect genuine feedback without being affected by other participants. These interviews were semi-structured. For blind participants' convenience, we invited them to stay in the main room, and each volunteer entered separate break-out rooms. When three researchers were available, each researcher interviewed one participant. If there were fewer than three, then one of the two researchers conducted one-on-one interviews with two participants sequentially.

During our interviews with *first-time participants* (new to paired-RSA), we followed up with the questions related to (1) exploring the opportunities and issues of paired-RSA; (2) understanding their transition from one-on-one RSA to paired-RSA; (3) interpreting their collaboration with other participants; and (4) probing potential (un)useful scenarios for paired-RSA. For *returning participants*, we collected their feedback on comparing sessions.

Besides interview questions, we asked blind participants to evaluate the usefulness of paired-RSA on a 5-point Likert scale (1 being the “least useful” and 5 being the “most useful”) for a session. The duration of interviews and collecting subjective feedback varied from 10 to 20 minutes.

Part 3: the focus group.—Lastly, we conducted a *focus group* with all the volunteers and blind participants in the main room to collect their feedback on their interactions with other participants. We prepared prompts to encourage dialog, including the (in)feasibility of paired-RSA and their collaboration with other participants. In the focus group, participants could evoke memories, share opinions, and have a consensus [56]. This part lasted for 5 to 10 minutes.

Data Analysis.—After the participants' consent, we recorded all sessions, including the interviews and focus groups. The first author manually transcribed the recorded data and analyzed the transcripts, using an iterative coding process with initial coding; then identified new concepts; and categorized them by themes and sub-themes [18]. All authors reviewed the concepts, themes, and sub-themes in weekly research meetings for months to finalize the codebook. Next, we present our findings.

5 FINDINGS

We found that paired-RSA is not only feasible but also desirable for certain tasks. In addition, we identified three distinct opportunities afforded by the rich collaboration in paired-RSA. In this section, we report these opportunities and analyze how paired-RSA works, and what issues may arise during the collaboration.

5.1 Paired-Volunteer RSA is Feasible

Overall, sixteen participants (V2-V8, U1-U9) appreciated *“the power of the second perspective”* (V6) brought by the second volunteer and indicated that paired-RSA could be beneficial in supporting complex and open-ended tasks, and enriching blind user’s quality of life and independence. A majority of participants (N = 7) found paired-RSA “most useful”; others considered paired-RSA as “useful” (N = 2) or “neutral” (N = 2). Next, we present several scenarios that reflect the desirability of paired-RSA and elaborate on the reasons, which answers RQ1 (distinctive opportunities offered by collaborative paired-RSA interactions).

5.1.1 Supporting Complex, Open-ended Tasks.—By adding a partner with opinions, sighted volunteers could coordinate their efforts to brainstorm, *“throw ideas”* (V5) and *“bounce things off each other”* (U9) to support blind participants with complex, open-ended tasks that require for multiple solutions or perspectives.

Thirteen participants (V2-V7, U2-U7, U9) gave positive feedback about applying paired-RSA to *“brainstorming”*, *“problem-solving”* and troubleshooting tasks, which not only offered more solutions for blind participants but also reduced the volunteers’ mental workload.

For instance, the tasks of learning the control panel on an electric grill and labeling microwave buttons. U8 requested assistance with labeling buttons on the microwave but lacked tactile stickers during the session. The blind participant and sighted volunteers, therefore, brainstormed about alternatives that were adhesive, available in U8’s home, and had distinct texture from that of the control panel. After proposing several ideas that were deprecated (e.g., tape, magnet), V7 run out of solutions. Fortunately, having *“another person with an opinion”* was helpful and efficient to *“brainstorm together what ways we could have that difference in texture”* (V7). The joint endeavors of volunteers contributed to the continuity of possible solutions for blind participants, which *“made things go a lot smoother and a lot quicker”* (U8).

U9 requested a similar task (labeling temperatures on a toaster) and echoed that paired-RSA is effective in troubleshooting tasks by allowing sighted volunteers to bounce ideas off each other with less mental strain.

“I think that when there [are] two of them together, they’re able to kind of have less mental strain and so they can think more flexibly and process better... I think they’d get flustered every once in a while... I’m sure like in the future if one got stuck, maybe the other would have some better idea, or you know, they can just kind of bounce things off each other.”

(U9)

Likewise, participants revealed the usefulness of paired-RSA in opinions-needed tasks with regard to providing more personal opinions for blind participants compared to regular RSA services. For example, applying makeup, matching outfits, and shopping online.

In opinions-needed tasks, sighted volunteers first delivered visual interpretation, e.g., “*automatically go through as far as sizes, colors, shapes, materials, prices*”(V6) for online shopping. Then, they gave subjective judgment based on blind participants’ requests. For example, U5 asked for volunteers’ opinions when choosing between red and pinkish lipsticks: “*Which one do you guys think would kind of fit... a great casual shirt on?*” and matching outfits: “*Hey, do you think this red matches?*”, “*Does this blue match?*”, “*Can these patterns go together?*”

Blind participants (U5, U7) reacted positively to paired-RSA in these open-ended tasks because they could obtain more opinions from different perspectives. U5 indicated that volunteers “*talked more*”, engaged more in expressing their subjective opinions than mere descriptions, which further enhanced her user experience of receiving more opinions about visual appeal.

“I did notice a couple of times they would give different descriptions or add to the other person’s description, especially when I asked their opinion. They did more versus just ‘What does this look like?’, like ‘Hey, does this outfit go together?’ I think both of them kind of talked more when it was ‘the opinion’ versus just describing something. So, that was cool. I think it is beneficial in that regard.”

(U5)

5.1.2 Enriching Quality of Life.—As RSA services have broadened in scope, blind users can now perform a set of high-stake daily tasks, such as navigating airports and shopping in large malls [50, 79]. However, being able to support RSA users to experience more of life in leisure activities and hobbies is also important and meaningful, which can enrich their quality of life. In paired-RSA, blind participants could receive assistance beyond basic interactions with the physical world and do more than tap around with their white cane. Paired volunteers complemented each other, entraining more detailed descriptions or easy-to-understand clarifications for blind participants in entertainment or pursuing hobbies. Nine participants (V2, V3, V6, V7, V8, U1, U3, U5, U6) indicated that paired-RSA is beneficial in “*subjective*”, “*descriptive*”, “*abstract*”, and “*imaginative*” tasks. For example, appreciating artwork, making origami, and describing multicolored yarn.

In session 1, U1 preferred to absorb as many verbal descriptions as possible for landscape paintings by Vincent Van Gogh, e.g., *The Mulberry Tree in Autumn* and *The Starry Night Over The Rhone*, which are abstract, emotional, and comprised of turbulent and various brushwork and a mysterious atmosphere. Even though being notified of and prepared for the task in advance, V3 said “*my stress levels immediately went up because I was like how in the world am I going to describe this?*”

By adding a second pair of eyes and a second perspective, volunteers felt confident and relaxed about having a “*backup*”, who “*came to rescue*” whenever they “*hit a wall with the description*” (V3) or “*run out of things to say*” (V2). V2 gave positive feedback about paired-RSA because the second perspective was “*effective*” in filling in the gap that she missed and “*mention[ing] things that go deeper*”. Likewise, V3 appreciated her collaboration with V2, where they complemented each other’s descriptions in turns by one articulating the color and feelings, and the other supplementing more details on brushstrokes and texture. Compared with regular RSA services provided by a single volunteer, the back-and-forth conversation between two volunteers evoked multiple perspectives and more details. Consequently, for the blind participant, paired-RSA enriched U1’s experience of artwork appreciation by gaining in-depth, comprehensive views of paintings and mapping them into her mind.

“I was able to get like two different viewpoints from each volunteer. So [like] if a volunteer didn’t pick up on one part of the painting, the other volunteer was able to pick up on it. And just overall be able to give a really in-depth view of what the paintings were.”

(U1)

Besides the complementary depictions or details, the second volunteer sometimes could provide more straightforward descriptions or clarifications for blind participants, particularly in “*tricky*” leisure activities, such as folding a paper airplane and making an angel with beads.

Even though visually obtained the content, sighted volunteers could have diverse ways to process the information and convert it into verbal instructions for blind participants. Some of the explanations successfully matched the blind participants’ understanding, while others did not. V6 commented that the second perspective mitigated his “*unconscious bias of being a sighted person*” when he took for granted that his descriptions of making a beaded angel were explicit and aligned with U3’s way of perception: “*Sighted people just take for granted that we can see all and interpret what we want or what we don’t want without being able to take it from [U3]’s perspective as far as what’s going to be important for [U3].*” As a recipient of visual interpretations, U2 corroborated that in the task of folding a paper airplane, the second perspective “*clicked with [his] brain*” when he couldn’t understand the other volunteer’s narration.

“There were a couple of times where I didn’t know exactly what the person was talking about, and where to fold the paper, what corner of the paper airplane. So it helps to have two people then because when one person got stumped and didn’t know how to describe, how to fold the paper, then the other person had a description that, for whatever reason, it clicked with my brain and it was easy [to understand] how they were describing it.”

(U2)

5.1.3 Promoting Independence of Blind Users.—Supported by call scheduling, blind participants could decide when and how to get assistance, and from whom to get

assistance. Put differently, paired-RSA can provide them more control over the utilization of RSA, promoting their agency and independence.

The aforementioned opportunities in paired-RSA extend this service into a broader coverage of more complicated, intellectual, experiential, and entertainment tasks with longer duration. Participants mentioned that it is less likely to accomplish these tasks in regular BME calls *“because it is a challenge to find someone to take the time to do that”* (U3). As an assistance provider, V6 elucidated that *“My average BeMyEyes call is 30 seconds to 3 minutes. That’s why anytime, anywhere I’ll always answer it, but I wouldn’t be able to do a 12-minute walkthrough on a pattern while I’m at work.”* Fortunately, by scheduling a mutual time, sighted volunteers (V1, V5, V6, V7) were willing to make a higher level of contribution to aid more time-consuming tasks, such as describing *“an entire movie”* or *“two-hour guided tour”*. This increased blind participants’ choices over the type of tasks they can get help with and for how long.

“If you could schedule, I think that would be kind of cool too, like how we did tonight with a longer session. If there was a way to say put a plea out and say, ‘Hey, would someone be willing to do, you know, a project with me for an hour? This is what I’m asking for.’ I think that would be really helpful too... I was able to ask good questions that I don’t think I would have been able to ask in regular BeMyEyes sessions because we were able to take the time.”

(U3)

Additionally, blind participants could specify their personal preferences by scheduling the call and raising their requests in advance. It promotes the feeling of independence by giving them more power in deciding how to approach a task. For example, blind participants could choose Van Gogh’s paintings in art appreciation (U1), audible or text-based instructions in navigation (U2), and even the gender of sighted volunteers based on their preference:

“Yeah, it would be really cool to be able to pick by [the] task because, I mean, some guys are good at fashion or makeup but probably not every guy. So, you know, I might feel more comfortable with a female.”

(U5)

5.2 Key Aspects of Paired-Volunteer RSA

In paired-RSA paradigm, cooperation between two volunteers and conversations between the volunteers and the blind participant contribute to new opportunities. In this section, we analyze how a paired-RSA session evolves and present three aspects that manifest effective collaboration among paired-RSA partners, which partially answer RQ2 (benefits of paired-RSA collaboration).

5.2.1 Creating Impromptu Roles among Sighted Volunteers.—In contrast to unidirectional assistance from the sighted volunteer to the blind user in one-on-one RSA, paired volunteers work collaboratively towards the goal of providing aid for the blind participant. Thus, a smooth collaboration between sighted volunteers is essential to the detailed, rich, insightful descriptions or opinions generated during paired-RSA. Sighted

volunteers created their roles spontaneously to better accommodate blind participants' requests. It promoted the volunteers' awareness concerning each other's responsibilities and expected actions, which influenced one's own decisions and readjusted their responsibilities.

Turn-taking is the strategy chosen by a majority of volunteers, where one volunteer was in full control of assistance for a period of time. The other volunteer in the "backseat" paid close attention to the conversation, prepared a mental list of things to add on, and avoid repetitive inputs.

"More kind of collaborative one where it turned into us just taking turns... I thought that was nice that having kind of just [V4] takes the lead. And that way I could observe and see, think about if there was anything I could add to help."

(V7)

Then, the volunteer who took the lead used verbal cues to remind the sighted partner of alternate (e.g., "*Hey, what do you think?*") or simply left pauses to ensure the occurrence of complementary descriptions or ideas for "*an equal opportunity to speak*" (V3).

Division of labor or parallelism is another strategy for coordination between volunteers, where both volunteers were devoted to specific subtasks at the same time. For example, V1 and V3 subdivided the task of helping U4 learn the functions of an electric grill. V3 described the layout of the control panel, streaming from U4's camera feed. Meanwhile, V1 searched online for the user manual of this appliance and supplemented step-by-step operations.

"Well, it's kind of nice because you can tag team. [V1] looked up the information and [V3] was just trying to describe the appliance in general, and then you know [V1] could look through it a little bit and find out more information to fill in what [V3] didn't know."

(U4)

5.2.2 More Engaging RSA Interactions.—Seven out of eight volunteers (except V8) shared their front-facing cameras, which helped the sighted peer to become more aware of their status. Both volunteers read each other facial expressions and communicated non-verbally by inferring the meaning of the peer's body language. Nods, smiles, raised eyebrows, and other non-verbal cues implied their acknowledgment, tiredness, and inquiry of turn-taking, constantly testing and confirming the formation of common ground. For example, V5 took her counterpart's action of drinking water as an implicit cue to take over: "*If I saw [V6] had really lengthy paragraphs that he was reading or he was taking like a sip of water, I took that as my cue to jump in because he obviously needed a little bit of a break.*"

Likewise, audio cues enhanced volunteers' awareness of their sighted partner's social context, and, more importantly, enabled the blind participants to hear volunteers' conversation and join in the interaction. Compared with regular RSA services, paired-RSA is "*putting multiple heads together, including the caller, you know, really becomes like a three-way conversation*" (V4). Ten participants (V2-V4, V6-V7, U1, U3, U5, U8-U9)

reacted positively to the social element embedded in paired-RSA interactions and described this three-way conversation as “cheerful”, “merrier”, “less awkward”, “fun”, “enjoyable”, “chatty”, and “a nicer, friendlier environment”, where both blind participants and sighted volunteers were comfortable, laughing and joking around. For instance, U9 appreciated how the conversation brought fun to the tedious chores rather than the zombies-like, objective instructions that she received in the task-oriented current RSA. V3 echoed that “human factor” in the three-way conversation distinguished the paired-RSA from and surpassed the regular one-on-one call:

“Like [U9] was saying, not just having someone like a zombie or like a robot saying instructions but having two human beings chatting with about your life or about your day, I think, can go a long way... It adds that human factor. Like it’s a group. It’s not just like a one-on-one thing that can get awkward or robotic.”

(V3)

5.2.3 Synchronized Displays to Establish Common Ground.—In regular RSA interactions, users stream the live video feed (i.e., their egocentric view) with the volunteers. However, this sharing is uni-directional as the users do not benefit from the camera feed of the volunteers. Holding the camera and pointing it to another digital display to share online content was inconvenient for both parties. For example, U2 felt fatigued or turned the camera in the wrong direction, and volunteers constantly guided U2 to adjust the angle of the camera back into the correct position. To improve the content-sharing strategy, the volunteers came up with two ideas on the fly: (1) the user could join the call from their computer and share the screens; (2) the user could share the link to online content, and one of the volunteers could share it with others via screen sharing. This idea was concurred by not only blind participants but also five volunteers (V1, V3-V4, V6-V7) who shared screens or links to artwork, tutorials, videos, or shopping websites with the group to create common ground directly. It pushed the boundary of regular RSA services, where users unidirectionally shared content with sighted volunteers. Through synchronized displays, volunteers could be more engaged in the RSA collaboration and inject more subjective opinions.

5.2.4 Ensuring Uninterrupted Assistance.—Having a second volunteer as a backup could maintain ongoing assistance if one of the volunteers encounters poor connectivity. U2 revealed that the current practice of RSA services under unstable Internet is to automatically lower the frame rate to a certain threshold and drop the calls if the connection degrades further. He gave an example where his call was dropped halfway through a destination; so he called again, “got someone else”, started over, and rebuilt the whole process by explaining his request and situation again in regular RSA services.

Experiencing the connection issue during Session 7, V6 indicated that the current practice is “100% redundancy for the sake of the user”. After collaborating with another volunteer, seven participants (V4, V6, V7, U1, U2, U8, U9) believed that paired-RSA could alleviate the poor-connectivity problem if it occurs on the volunteer’s side.

V4 had connectivity issues when assisting U8 with reading names and describing images on cases of vinyl records, *“Lawrence Welk and his champagne... [disconnected]”*. Meanwhile, the paired volunteer (V7) became aware of the circumstances from V4’s choppy video and distorted audio, and reacted quickly by continuing the description, *“they’re playing those like accordions”*. Afterward, V4 pointed out that the help of the second volunteer *“definitely made it a lot less stressful to drop out because I knew that it would continue without me”*. U8 also had a positive response to having a backup when V4 dropped out unexpectedly, as it got rid of the lengthy process of calling back and starting over.

“You’re not losing out on the call, you’ve got, you know, other resources, and it’s easy enough to get back into the call. [V4] dropped out like twice, but I could still understand what she was trying to say because after [she] came back, it retained what she was saying...”

(U8)

5.3 Issues of Paired-Volunteer RSA

Although a rough framework for coordination was negotiated early on in the session, task demands, communication demands, and group members’ commitment or skill might stray from what they planned or expected in the course of cooperation. The divergence could undermine collaboration and culminate in a breakdown. In this section, we present several issues in paired-RSA and analyze the reasons, which partially answer RQ2 (cost of paired-RSA collaboration).

5.3.1 Diversity and Conflict.—Participants expressed apprehensions about miscommunications in a general context — individual differences leading to tension or group cohesion problems, although these occurred rarely in our study.

Personality difference between volunteers is one example. Seven participants (V1, V2, V5, V7, U4, U5, U9) mentioned that *“power and control type things”* (V1) and chaos might occur if mismatching volunteers’ personalities. They were worried about assigning *“two very opinionated people”* (V7) in paired-RSA, or volunteer running into a counterpart who is stubborn, ego-centric, *“aggressive and likes to take over the call”* (V2). Tensions between volunteers could hinder the assistance and further impact blind participants’ experience, as U9 explained, *“I think that if the agents didn’t get along with each other if they had tension, you would feel it in the room. You wouldn’t want that.”*

Conflicting styles of assistance and gaps in expertise are two more reasons why collaboration could break. Due to the large number of volunteers registered on RSA platforms, the distribution of knowledge and skill across volunteers is not uniform. For example, volunteers’ knowledge about the world (education background), vision impairments, and orientation and mobility training is likely to vary. In these cases, V1 recommended post-assistance training rather than training novices and correcting their mistakes during paired-RSA.

V1 expounded his point of view from two aspects. Firstly, as an experienced volunteer, V1 found himself *“having to bite my tongue a lot”* and *“kind of sit back”* to give novices

more opportunities to practice, although they may perform less suitably or professionally. Therefore, V1 struggled between being silent to allow novices to practice more and interjecting to ensure appropriate assistance for blind participants: “[It] seemed very challenging because like, I didn’t want to like step on [V3]’s toes. But at the same time, I’m like, I kind of feel you’re going down the wrong path.”

Secondly, for the sake of users, it is less efficient if volunteers devote more effort to adjusting, correcting, and training as opposed to getting the task done. Conflicting styles between novices and experts could also confuse blind participants, as V1 explained, “I think that it gets really confusing when we’ve got these two kinds of differing ways of accomplishing something getting thrown at the person that needs the assistance, so I think that it’s almost problematic.”

5.3.2 Boundary Conditions.—Coordination breakdown occurred in “objective”, “simple”, “minor” and “fairly easy” tasks, such as reading mail, a newspaper, or labels of medicine or cans. After familiarizing themselves with tasks and creating roles, volunteers had presuppositions about what to do and how much to contribute during the session. However, the demand for assistance with objective tasks was lower than their presupposition because these tasks sometimes were and could be completed by just one volunteer. Therefore, the input from the second volunteer was dispensable, as V3 put it, “whatever you are reading or narrating for the user can’t be disputed”. It culminated in a breakdown that role assignments or even the attendance of the second volunteer were unnecessary.

Five participants (V2, V3, V5, V7, U1) considered their or their counterpart’s commitment was less than expected; thus, they believed that paired-RSA was less beneficial in this scenario. Three participants (V1, V3, U1) were even concerned that paired-RSA for minor, objective tasks increased labor cost by “taking up that second [sighted] person’s time [who] could be helping another person doing a different task” (U1). To address this concern, participants suggested adding an option to choose regular RSA or paired-RSA (U1), or adding a second volunteer halfway through the call if needed (V3, V5, U2, U5).

5.3.3 Information Overload.—Tasks that involved navigational instructions were another example of cooperation breakdown for two reasons. First, role assignments were complex in this highly-dynamic context, with one overlapping responsibility shared among the volunteers — ensuring blind participants’ safety. Navigational tasks entail several uncertainties and risks, such as obstacles, a narrow view of the camera, and delayed or decoupled information on location or orientation. Volunteers were well aware of these issues centered on navigational tasks; thus, both of them prioritized blind participants’ safety and alerted verbally and immediately when risks emerged, before they got a consensus on how to warn the blind participants.

Second, navigational tasks are time-sensitive, “fast-peaced” (U6), where blind participants appreciate “quick, directed and uninterrupted narration” (V4). It is evident by V1’s professional RSA experience: when he continuously asked blind users, “Can I please have you stop so I can look at the map? Can I have you stop so I can look up this? Can I have... They don’t want to stop.” Due to blind participants’ preference for a succinct stream of

rapid instructions, volunteers felt temporal demand and delivered instructions directly before reaching a consensus with each other. Consequently, paired-RSA “*expands or increases the information that’s actually happening*” (V4), overwhelms both blind participants and sighted volunteers with either conflicting or repetitive instructions. For instance, without establishing common ground, V1 and V4 provided opposite directions to U6 when assisting her in grocery shopping.

“They were directing me way to put my camera, and I don’t think they were quite in sync with each other, because just as you know, one volunteer said, ‘Oh, put it to the right,’ and the other was like ‘Actually, you know, I think I can see if you put it to the left’. There were some situations like that, so I did have to move my camera around a lot.”

(U6)

Likewise, five participants (V2, V4, V7, U6, U8) expressed similar concerns that paired-RSA could generate conflicting instructions that confuse blind participants in navigation. Seven participants (V1, V4, V7, V8, U2, U3, U5) were also worried that even if instructions from two volunteers are consistent and do not pull blind participants in different directions, they could be verbose, overloading the audio channel and further confusing the users.

To remediate the common ground, five participants (V1, V8, U1, U2, U4) suggested assignments of more well-defined roles, as well as parallelism during paired-RSA. In the challenging, time-pressured navigational tasks, volunteers need to constantly check the live video feed to ensure blind participants’ safety, meanwhile, plan the path on maps to create a seamless experience for them. Paired-RSA could distribute the cognitive load among the volunteers through the division of labor, and thus potentially reduce the amount of information processed by each volunteer. V8 elucidated that one volunteer could play the role of “*silent partner*” who supports the other volunteer by searching maps online, figuring out the layout of premises, and planning a route to the blind participant’s desired destination, but not delivering information to the blind participant. In the meantime, the other volunteer could focus on the “*customer service*”, and provide assistance seamlessly by verbally navigating the blind participant without pausing.

“I will say that in the realm of navigation, it may be helpful to have a silent partner, and what I mean by that is that somebody that’s pulling up the map and doing kind of the groundwork ahead of time, or while the user is getting set up. So that one agent can focus on the customer service and not pausing while helping the user, and the other person helping the user can be bringing up the maps to help navigate.”

(V8)

6 DISCUSSION

In this section, we synthesize the taxonomy of tasks where paired-RSA can succeed and identify new opportunities in this new RSA paradigm, which is extending to N-volunteers RSA and utilizing algorithmic decision-making for pair matching. Finally, we present the limitations and the directions for future work.

6.1 Taxonomy of Tasks

We present the taxonomy of tasks in Table 5. Our taxonomy contains five dimensions:

- D1** whether the task *types* are based on personal interpretations or opinions (e.g., subjective, objective, or both);
- D2** what *content is shared* (e.g., users' live camera feed, users' device screen showing visual content, volunteers' device screen showing visual content);
- D3** the amount of *temporal and cognitive demand* involved in completing the task (high, low);
- D4** whether *special skills* are needed (yes, no); and
- D5** whether *labor can be divided*, allowing sighted volunteers to perform subtasks in a serial, parallel, or hybrid manner.

Three of these dimensions – D1: type, D3: temporal and cognitive demand, and D4: the need for special skills – are intrinsic to tasks and remain consistent across both regular RSA and paired-RSA services. However, dimension D5: division of labor (serial, parallel, or hybrid) is unique to paired-RSA interactions, which emerges organically during collaborations between multiple volunteers (Section 5.2.1). Similarly, dimension D2: content sharing through synchronized displays – either screen or live camera feed – emerged from our study as an extension of RSA that enables more intellectual and experiential tasks. By allowing both blind participants and sighted volunteers to share and analyze a common referent (e.g., online videos, and shopping websites), blind users can receive assistance beyond basic interactions with their immediate physical environment (e.g., reading, navigation).

6.2 Comparison with Generic Frameworks for Collaborative Work

Compared to McGrath's group tasks circumplex—Compared to McGrath's group tasks circumplex, described in Section 2.3, our taxonomy isn't mutually exclusive – that is, a task isn't confined to one dimension. However, it is collectively exhaustive, meaning all tasks fit within some dimension. This taxonomy proves useful as it highlights differences and relations among tasks that might otherwise go unnoticed. Tasks with the attribute value “objective” in our dimension D1: Type broadly fall under McGrath's group task category 2, which involves choosing a solution or plan from a set of alternatives where a correct or agreed-upon answer exists. Conversely, tasks with the attribute value “subjective” fall under McGrath's group task category 1, which involves generating ideas or plans. Tasks that can be either objective or subjective in our dimension D1: Type and require a high temporal and cognitive demand (D3) or special skills (D4) are likely to fall under McGrath's group task category 2, which involves negotiation to reach consensus, or category 4, which involves executing a plan.

However, these relationships are nuanced, as our exhaustive permutation of five dimensions yields a large number of task configurations ($3 * 3 * 2 * 2 * 3 = 108$); many of these may not neatly align with McGrath's four group task categories. We leave this exploration for future work.

Compared to Bales’s group interaction processes—Compared to Bales’s group interaction processes, also described in Section 2.3, our findings differ significantly as we did not observe a consistent “talker” who coordinates the tasks [17]. Typically, this talker is well-positioned (e.g., the leader) in the group to initiate and address most of the communications [41, 74, 75]. A potential reason for this is the lack of assigned roles or leadership in paired-RSA; sighted volunteers create their roles spontaneously to better accommodate the requests of blind participants.

However, we did observe other aspects of group interaction processes, such as actively providing suggestions and options or passively seeking these inputs, along with a mix of positive and negative expressions to either release or create tension or to convey agreement or disagreement. We found that these interactions effectively reduced individual volunteer bias (i.e., averaged out [55]), making paired-RSA particularly suitable for subjective tasks.

6.3 Potential Success Areas for Paired-RSA

Based on the findings, we identified that paired-RSA could succeed in subjective tasks; objective tasks that require specialized skills, or a high level of cognitive load; and potentially benefit tasks where parallelism could be applied. Next, we elaborate on the usefulness of paired-RSA in these categories.

Multiple perspectives were appreciated by both blind participants and sighted volunteers in subjective tasks with open-ended solutions. First, multiple volunteers could provide blind participants with diverse, in-depth, and comprehensive descriptions of aesthetic content that is abstract, emotional, and imaginative, such as paintings and patterns of jewelry. Second, multiple perspectives could mitigate volunteers’ “*unconscious bias of being a sighted person*” (V6), refine and clarify the other volunteer’s narrations for blind participants’ better understanding of tasks with complicated steps (e.g., crafting). Third, sighted volunteers could bounce suggestions on opinions-needed tasks (e.g., online shopping and fashion help) and offer more personal opinions if required.

Paired-RSA could also smooth out assistance on complex objective tasks that require specialized skills or a high level of cognitive load. For instance, help with household appliances and reading multilingual text. Sighted volunteers brainstormed solutions to unfamiliar problems (e.g., labeling microwave buttons) so they could have little “downtime” and become less stressed. Meanwhile, blind participants could receive continuous assistance if one of the volunteers runs out of solutions. When assisting blind participants with tasks that require topical knowledge (e.g., different languages), multiple volunteers could bring a variety of expertise and complement each other. For example, we paired a volunteer who can read Korean and a volunteer who can read Spanish in Session 1 (Table 4). As such, blind participants could have a seamless experience by receiving help from various experts all together in paired-RSA, rather than calling regular, one-on-one RSA multiple times and accomplishing piece by piece.

6.4 Potential Limitations of Paired-RSA

Our findings revealed that paired-RSA was less beneficial in directional, time-pressured, high-stake tasks where instructions might be very sporadic, and spur of the moment based on the condition of blind participants' physical surroundings. For instance, grocery shopping and navigation tasks. To ensure blind participants' safety and satisfy their preference for quick instructions, volunteers could deliver conflicting instructions related to the directions, or lead blind participants in the same direction but with different expressions before reaching a consensus. However, if volunteers subdivide the task and execute it in parallel, paired-RSA could be helpful for these multi-thread tasks by reducing the workload of each volunteer and contributing to a seamless experience for blind participants. During navigation, one volunteer could play the role of *"a silent partner"* who plans ahead but does not involve in delivering information to blind participants. Meanwhile, the other volunteer could *"focus on the customer service and not pausing while helping the user"* (V8).

Likewise, if paired-RSA is utilized for training purposes, the experienced volunteer could be a silent partner, observing the novice's performance during assistance and offering advice afterward. As indicated by V1, when pairing experienced volunteers with novices, the gaps in expertise could bother the former by struggling between keeping silent and interjecting, and also situate blind participants in tension between volunteers (Section 5.3.1). To alleviate this problem, the experienced volunteer could take on the role of a silent partner and give the novice more opportunities to practice in real scenarios involving blind participants. Besides, the post-assistance training could allocate more mentoring time to volunteers than immediately adjusting or correcting novices' performance.

The idea of parallelism can also be applied to, and merit other scenarios where external resources are needed. For example, one volunteer can search for manuals of appliances in trouble-shooting tasks or details of products in grocery shopping, while the other volunteer monitors the live camera feed and endeavors into verbal instructions.

6.5 Extending Volunteer-based RSA to N-volunteers RSA

We proposed the design idea of pairing a user with multiple (N) volunteers in a call to increase the utilization of volunteering resources and promote the social good of understanding blind people. In this paper, we carried out an exploratory study to specifically investigate the scenario where N equals 2. By bringing in a second-sighted person, N-volunteers RSA creates needs that are not present in the original RSA paradigm. Sighted volunteers need to coordinate with each other and the user through verbal and non-verbal communication, as opposed to merely cooperating with a blind user. In this section, we first categorize the differences in characteristics of RSA and N-volunteers RSA, then examine the feasibility of N-volunteers RSA through indicators of success in remote collaboration defined by Olson and Olson [59].

Our study revealed that the proposed paired-RSA transforms the characteristics of traditional RSA as follows:

1. ***IvI***: paired-RSA transfers *1-on-1* communication to *1-on-N* ($N = 2$) communication, providing opportunities for cooperation in synchronous interactions.
2. ***bi-directional audio***: paired-RSA extends the two-way audio conversation to three-way audio-video hybrid-channel communication (e.g., the volunteers can engage in a video-chat-like experience).
3. ***unidirectional video***: paired-RSA augments unidirectional back-facing camera feed sharing of RSA to a more complex video-based sharing paradigm: *i*) The two volunteers can also share their front-facing live camera feed for better cooperation. Note that this setup is not necessary for traditional RSA, since blind users may not perceive the visual content. *ii*) All three persons in a session (the user and two volunteers) are able to share non-camera feed content (e.g., computer screen) to establish common ground.
4. ***objective tasks***: paired-RSA can not only be used for supporting objective tasks but also subjective tasks, where paired volunteers can provide checks and balances for each other's opinions and reach common ground, thus reducing individual biases.

Four indicators of successful remote collaboration defined by Olson and Olson [59] attribute to this smooth transition from RSA to N-volunteers RSA. First, *common ground* [26] refers to the knowledge that people have in common and are aware that they have that knowledge in common. In this study, participants established and maintained the common ground by familiarizing themselves with tasks prior (awareness of what to do or what to expect), creating roles spontaneously on the fly (volunteers knowing when to do what), sharing non-camera feed content bi-directionally (awareness of what are collaborators referring to), and synchronizing social context (awareness of collaborators' status).

Second, *coupling*, which refers to the scope and type of communication, is required in teamwork. In the original RSA setup, the communication is a two-way conversation between the blind user and sighted volunteer, and uni-directional video streaming from the user to the volunteer. In contrast, the communication during paired-RSA is three-way, social, and conversational in nature over the audio-video channel. It enhances the common ground, particularly the user and volunteers could be better aware of their partners' status (e.g., who is talking? what are my partners doing?).

Third, *collaboration readiness*, which means the underlying culture of sharing and collaboration among partners. RSA is a full-fledged collaborative infrastructure, where blind users request help with daily tasks from sighted volunteers and collaborate on tasks. Both parties embrace, engage, and appreciate the culture of cooperation. Taking BME for example, there are 4.5 million blind and low-vision individuals signed up worldwide as users and over 6 million volunteers, indicating both parties' willingness to participate in remote collaboration. Additionally, the adequate volunteer resource on BME makes the idea of N-volunteers RSA feasible. With the disproportionate ratio of blind users to sighted volunteers (1:13), we could explore more possibilities in N-volunteers RSA, where N could be 2 or greater.

Fourth, *technology readiness*. RSA services are supported with mature technologies, which can be operated and maintained on a large user basis. RSA platforms have been developed and tested in both academic research (e.g., VizWiz [15], BeSpecular [40]) and technology industry (e.g., BeMyEyes [6]). In terms of the future development for paired-RSA, we found the currently-available video chat platforms (e.g., Zoom) could be a technical approximation and satisfy the technology demand because it contains all the functions available on RSA platforms (e.g., two-way audio connection, video transmission, screenshot) and even those that are not supported in RSA platforms (e.g., a multitude of camera angles, screen sharing). This is evident by the feasibility of three-way conversations in the study and by no technology glitches reported by participants.

Although quantifying the performance of volunteers and the benefit of N-volunteers RSA is meaningful and achievable [55], it is beyond the scope of this work. For time-sensitive tasks (e.g., navigating in grocery stores), performance improvement from RSA to N-volunteers RSA is a primary manifestation of effectiveness. However, in contrast, it is a less critical assessment in open-ended, entertainment tasks (e.g., making origami, appreciating artwork, matching outfits). Instead, we aim at emphasizing humanity and enhancing the overall enjoyment level, where blind and sighted people encounter others, work meaningfully together, and understand people who are differently sighted. In sum, this exploratory study opens up a possibility for more studies to investigate more details of N-volunteers RSA and delve deeper into other complex, open-ended tasks in the future.

6.6 Algorithmic Decision-Making for Pair Matching

Our study revealed the difficulty of finding a perfect match in paired-RSA and identified two factors that could affect performance: skill matching and personality matching. The randomness in user-volunteer pairing is similar to information asymmetries [67] and opaque assignment managements [54] experienced by Uber drivers, where they have less than 15 seconds to decide whether to accept or reject the request without knowing its content (e.g., destination and pickup location). Consequently, drivers accept the ride request and cancel it when they find out the cost of reaching the pickup location is higher than the remuneration. This confusion and opaqueness lead to negative impacts on both drivers and passengers.

Similarly, blind users on volunteer-based RSA platforms could become frustrated repeatedly making calls for expertise-required tasks to connect with sporadic volunteers who have pertinent skills among a large pool of registered volunteers. To address this randomness, we scheduled calls and mapped skills in this study. This process was completed transparently and manually, where we collected users' requests, sent them to volunteers, and inquired about volunteers' willingness to help and whether they possess pertinent expertise. Regardless of the benefits or uncertainties added by paired-RSA, we received no negative responses to skill mapping, implying the desirability of this design idea and laying the foundation for a future, transparent, AI-driven, and large-scale skill-matching system that could be integrated into RSA services.

Personality matching was not controlled in this study, but it is important to mitigate miscoordination as indicated by participants (Section 5.3.1). For example, not pairing very opinionated or reticent people. Personality differences between volunteers could incur

tensions and degrade the user experience for both sighted volunteers and blind users. In contrast, matching partners' personalities could better leverage the trait of social engagement embedded in paired-RSA and make the interactions joyful for both sides.

We envision that users' descriptions of request types, volunteers' preferences for tasks or skill sets, and their patterns of personalities are essential in the automatic, transparent pair-matching system. This information could be saved into blind users' and sighted volunteers' profiles. On the one hand, users could save their frequent tasks and details to the profiles, such as fashion styles, types, and brands of appliances. On the other hand, volunteers could identify their skill sets, including technology, cooking, sports, fashion, and O&M training. Besides, volunteers could choose their (un)willingness to be paired with specific personality traits and their preferred roles (e.g., communicator or silent partner) to avoid conflicting personalities and tensions. With the association with profiles and a transparent pair-matching system, both parties would be more efficient and desirable in terms of coordination and tasks that require topic-specific expertise. The implementation of such a system is straightforward that similar skill-matching systems have been widely used in job search engines [65] or professional social matching (PSM) [60] using natural language processing (NLP) and machine learning (ML) algorithms.

6.7 Comparison with Multiple Captions in Image Captioning and VQA

For the tasks of describing video contents, our idea of paired-RSA shares some similarities with multiple captions in image captioning and visual question answering (VQA). The AI community has recently developed machine learning models to automatically create captions for images or answer visual questions, especially for that originating from real users such as blind people [36, 37]. The algorithms are trained from large-scale crowdsourcing labeled datasets [71], and the datasets usually collect multiple labels for each image or visual question. For example, the VizWiz-VQA dataset [36] collected 10 answers per visual question; the VizWiz-Captions dataset [37] collected five captions for each image. Bhattacharya *et al.* [14] identified nine plausible reasons (e.g., subjective, ambiguous) why different people provide different answers in annotating the datasets and found that the algorithms could generate better answers by synthesizing multiple persons' perceptions of images and language. In line with this finding, our exploratory empirical study also found that multiple volunteers could provide blind users with complementary descriptions or easy-to-understand clarifications using a turn-taking strategy.

Despite the similarity, our paired-RSA paradigm is substantially different from multiple captions in the creation of VQA datasets. First, multiple captions in VQA datasets were asynchronously, and parallelly collected from different crowd workers, but paired-RSA is a synchronized, collaborative service via a live video conversation between two sighted volunteers and a blind user. The paired-RSA involves complex dynamic interactions when describing video contents. Second, multiple captions in VQA are used for developing an AI-based support system, but RSA is a human-driven service. Our study on paired-RSA is purely human-centered, and we are expanding RSA and emphasizing more human-human interactions among three people. We will leave the study on the interactions between multiple humans and AI [51] as future work.

6.8 Limitations and Future Work

We tried to pursue a positive opportunity of extending RSA and making the coverage of the current RSA technique broader. We carried out an exploratory study to explore the feasibility of performing subjective, entertainment, specialized skills-required, high-temporal demand, or high-stake tasks in paired-RSA. Tasks completed in this study were regarded as typical by the informants and were chosen by blind participants. Still, we examined a relatively small set of real-world tasks that users frequently need help with. This limitation is inevitable in early-stage, exploratory research, thus broadening the types of tasks is an important future trajectory.

We acknowledge that pairing two volunteers is a subset of N-volunteers RSA, which is simple and applicable. Situating RSA in a larger group is the potential to better leverage a resource of volunteers' perspectives, skills, and insights, but it could impose more uncertainties and complexities. One issue is the tradeoff between the amount of information provided and the partners' ability to comprehend and utilize it. Volunteers need to check live video feeds from their sighted partners and blind user, deliver descriptions and make adjustments as needed, prepare for turn-taking, and sometimes refer to external resources. Similarly, the blind user needs to recognize who is talking and what is the content and could be overwhelmed by conflicting or repetitive instructions. Therefore, our future direction is to pair more than two volunteers and explore the balance between the amount of information for attaining group consensus and avoiding information overload.

7 CONCLUSION

In this paper, we delve into RSA services by proposing two design ideas to better leverage the huge resource of volunteers and extend the coverage of services: pairing a user with two volunteers in a call and supporting call scheduling. To investigate the feasibility of the design ideas, we designed a variety of tasks for paired-RSA and call-scheduling setting and tested with 9 blind users and 8 sighted volunteers. The results show that the two design ideas are favored by both parties and demonstrate the potential enhancement for RSA services: (1) *entertainment* – helping users pursue hobbies and enrich leisure activities; (2) *humanism* – encouraging subjective answers in opinions-needed tasks instead of mechanistic objective descriptions; (3) *efficiency* – enabling parallel assistance in troubleshooting tasks; (4) *stability* – introducing a backup in case of low-bandwidth conditions; (5) *quality* – enhancing volunteers' assistance skills when collaborating with other volunteers; (6) *sociality* – increasing social engagement in the three-way conversation. Based on these potential benefits, we can envision the features of volunteer-pairing and call-scheduling incorporated into the next iteration of RSA paradigm.

ACKNOWLEDGMENTS

We thank *Be My Eyes* and Penn State Student Disability Resources Office for their cooperation in this work. We also thank the anonymous reviewers for their insightful comments and Emanuel Mendiola-Ortiz for his contribution to this work. This research was supported by the US National Institutes of Health, National Library of Medicine (R01 LM013330).

REFERENCES

- [1]. 2022. 100 Ways to Use Be My Eyes. <https://support.bemyeyes.com/hc/en-us/articles/360005528317-100-Ways-to-Use-Be-My-Eyes>.
- [2]. 2022. Aira. <https://aira.io/>.
- [3]. 2022. Autour. <http://autour.mcgill.ca/en/>.
- [4]. 2022. Be My Eyes – Specialized Help. <https://www.bemyeyes.com/specialized-help>.
- [5]. 2022. Be My Eyes Help Center – Why am I not receiving any calls? <https://support.bemyeyes.com/hc/en-us/articles/360005894637-Why-am-I-not-receiving-any-calls->.
- [6]. 2022. Bringing sight to blind and low-vision people. <https://www.bemyeyes.com/>.
- [7]. 2022. Can I change my availability hours? - Be My Eyes Help Center. Retrieved February 12, 2022 from <https://support.bemyeyes.com/hc/en-us/articles/360005892797-Can-I-change-my-availability-hours->
- [8]. 2022. Getting Started with Be My Eyes. <https://support.bemyeyes.com/hc/en-us/articles/360005536558-Getting-Started-with-Be-My-Eyes>.
- [9]. 2022. The Seeing Eye GPS™ App in the iTunes Apple Store! <http://www.senderogroup.com/products/shopseeingeyegps.html>.
- [10]. 2022. Specialized Help – Be My Eyes Help Center. <https://support.bemyeyes.com/hc/en-us/articles/360012867978-Specialized-Help>.
- [11]. 2022. What to Expect from Be My Eyes Calls. <https://support.bemyeyes.com/hc/en-us/articles/360005542717-What-to-Expect-from-Be-My-Eyes-Calls>.
- [12]. Avila Mauro, Wolf Katrin, Brock Anke, and Henze Niels. 2016. Remote assistance for blind users in daily life: A survey about Be My Eyes. In Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments. 1–2.
- [13]. Baranski Przemyslaw and Strumillo Pawel. 2015. Field trials of a teleassistance system for the visually impaired. In 2015 8th International Conference on Human System Interaction (HSI). IEEE, 173–179.
- [14]. Bhattacharya Nilavra, Li Qing, and Gurari Danna. 2019. Why does a visual question have different answers?. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 4271–4280.
- [15]. Bigham Jeffrey P, Jayant Chandrika, Miller Andrew, White Brandyn, and Yeh Tom. 2010. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, 65–72.
- [16]. BlindSquare. 2020. BlindSquare iOS Application. <https://www.blindsquare.com/>.
- [17]. Borgatta Edgar F and Bales Robert F. 1953. Interaction of individuals in reconstituted groups. *Sociometry* 16, 4 (1953), 302–320.
- [18]. Bryman A and Burgess RG. 1994. Analyzing Qualitative Data. Routledge. <https://books.google.com/books?id=KQkotSd9YWkC>
- [19]. Bujacz M, Baranski P, Moranski M, Strumillo P, and Materka A. 2008. Remote guidance for the blind—A proposed teleassistance system and navigation trials. In 2008 Conference on Human System Interactions. IEEE, 888–892.
- [20]. Burton Michele A, Brady Erin, Brewer Robin, Neylan Callie, Bigham Jeffrey P, and Hurst Amy. 2012. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. In Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility. ACM, 135–142.
- [21]. Carroll John M., Lee Sooyeon, Reddie Madison, Beck Jordan, and Rosson Mary Beth. 2020. Human-Computer Synergies in Prosthetic Interactions. *IxD&A* 44 (2020), 29–52. http://www.mifav.uniroma2.it/inevent/events/idea2010/doc/44_2.pdf
- [22]. Carroll John M, Neale Dennis C, Isenhour Philip L, Rosson Mary Beth, and McCrickard D Scott. 2003. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies* 58, 5 (2003), 605–632.

- [23]. Carroll John M, Shih Patrick C, Han Kyungsk, and Kropczynski Jess. 2017. Coordinating community cooperation: Integrating timebanks and nonprofit volunteering by design. *International Journal of Design* 11, 1 (2017).
- [24]. Chaudary Babar, Paajala Iikka, Keino Eliud, and Pulli Petri. 2017. Tele-guidance based navigation system for the visually impaired and blind persons. In *eHealth 360*. Springer, 9–16.
- [25]. Chaudary Babar, Pohjolainen Sami, Aziz Saima, Arhippainen Leena, and Pulli Petri. 2021. Teleguidance-based remote navigation assistance for visually impaired and blind people—usability and user experience. *Virtual Reality* (2021), 1–18.
- [26]. Clark Herbert H. 1996. *Using language*. Cambridge university press.
- [27]. Das Maitraye, Gergle Darren, and Piper Anne Marie. 2023. *Simphony: Enhancing Accessible Pattern Design Practices among Blind Weavers*. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [28]. Fallah Navid, Apostolopoulos Ilias, Bekris Kostas, and Folmer Eelke. 2013. Indoor human navigation systems: A survey. *Interacting with Computers* 25, 1 (2013), 21–33.
- [29]. Fussell Susan R, Kraut Robert E, and Siegel Jane. 2000. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 21–30.
- [30]. Fussell Susan R, Setlock Leslie D, and Kraut Robert E. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 513–520.
- [31]. Garaj Vanja, Jirawimut Rommanee, Ptasinski Piotr, Cecelja Franjo, and Balachandran Wamadeva. 2003. A system for remote sighted guidance of visually impaired pedestrians. *British Journal of Visual Impairment* 21, 2 (2003), 55–63.
- [32]. Greer Jim E, Mccalla Gordon, Collins Jason A, Kumar Vive S, Meagher Paul, and Vassileva Julita. 1998. Supporting peer help and collaboration in distributed work-place environments. *International Journal of Artificial Intelligence in Education (IJAIED)* 9 (1998), 159–177.
- [33]. Grice Herbert P. 1975. *Logic and conversation*. In *Speech acts*. Brill, 41–58.
- [34]. Grice H Paul. 1978. Further notes on logic and conversation. In *Pragmatics*. Brill, 113–127.
- [35]. Guerreiro João, Sato Daisuke, Asakawa Saki, Dong Huixu, Kitani Kris M, and Asakawa Chieko. 2019. CaBot: Designing and evaluating an autonomous navigation robot for blind people. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 68–82.
- [36]. Gurari Danna, Li Qing, Stangl Abigale J, Guo Anhong, Lin Chi, Grauman Kristen, Luo Jiebo, and Bigham Jeffrey P. 2018. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3608–3617.
- [37]. Gurari Danna, Zhao Yanan, Zhang Meng, and Bhattacharya Nilavra. 2020. Captioning images taken by people who are blind. In *European Conference on Computer Vision*. 417–434.
- [38]. Hackman J Richard. 1968. Effects of task characteristics on group products. *Journal of Experimental Social Psychology* 4, 2 (1968), 162–187.
- [39]. Holmes Nicole and Prentice Kelly. 2015. iPhone video link facetime as an orientation tool: remote O&M for people with vision impairment. *International Journal of Orientation & Mobility* 7, 1 (2015), 60–68.
- [40]. Holton Bill. 2016. BeSpecular: A new remote assistant service. *Access World Magazine* 17, 7 (2016).
- [41]. Horvath William J. 1965. A mathematical model of participation in small group discussions. *Behavioral Science* 10, 2 (1965), 164–166. [PubMed: 14284291]
- [42]. Jain Dhruv. 2014. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. 225–232.
- [43]. Johnston Michael V.. 2009. Plasticity in the developing brain: Implications for rehabilitation. *Developmental Disabilities Research Reviews* 15, 2 (2009), 94–101. 10.1002/ddrr.64arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ddrr.64 [PubMed: 19489084]
- [44]. Kamikubo Rie, Kato Naoya, Higuchi Keita, Yonetani Ryo, and Sato Yoichi. 2020. Support strategies for remote guides in assisting people with visual impairments for effective indoor

navigation. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 1–12.

- [45]. Kraut Robert E, Fussell Susan R, and Siegel Jane. 2003. Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction* 18, 1–2 (2003), 13–49.
- [46]. Kutiyawala Aliasgar, Kulyukin Vladimir, and Nicholson John. 2011. Teleassistance in accessible shopping for the blind. In Proceedings on the International Conference on Internet Computing (ICOMP). The Steering Committee of The World Congress in Computer Science, Computer ..., 1.
- [47]. Lasecki Walter S, Wesley Rachel, Nichols Jeffrey, Kulkarni Anand, Allen James F, and Bigham Jeffrey P. 2013. Chorus: a crowd-powered conversational assistant. In Proceedings of the 26th annual ACM symposium on User interface software and technology. ACM, 151–162.
- [48]. Laughlin Patrick R. 2015. Social combination processes of cooperative problem-solving groups on verbal intellectual tasks. In *Progress in social psychology*. Psychology Press, 127–155.
- [49]. Lee Sooyeon, Reddie Madison, Gurdasani Krish, Wang Xiyang, Beck Jordan, Rosson Mary Beth, and Carroll John M.. 2018. Conversations for Vision: Remote Sighted Assistants Helping People with Visual Impairments. arXiv:1812.00148 [cs.HC]
- [50]. Lee Sooyeon, Reddie Madison, Tsai Chun-Hua, Beck Jordan, Rosson Mary Beth, and Carroll John M. 2020. The emerging professional practice of remote sighted assistance for people with visual impairments. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 1–12.
- [51]. Lee Sooyeon, Yu Rui, Xie Jingyi, Billah Syed Masum, and Carroll John M. 2022. Opportunities for human-AI collaboration in remote sighted assistance. In 27th International Conference on Intelligent User Interfaces. 63–78.
- [52]. Li Bing, Munoz J Pablo, Rong Xuejian, Xiao Jizhong, Tian Yingli, and Arditi Aries. 2016. ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In European Conference on Computer Vision. Springer, 448–462.
- [53]. Loomis Jack M, Klatzky Roberta L, Golledge Reginald G, et al. 2001. Navigating without vision: basic and applied research. *Optometry and vision science* 78, 5 (2001), 282–289. [PubMed: 11384005]
- [54]. Ma Ning F, Yuan Chien Wen, Ghafurian Moojan, and Hanrahan Benjamin V. 2018. Using stakeholder theory to examine drivers’ stake in Uber. In Proceedings of the 2018 CHI conference on human factors in computing systems. 1–12.
- [55]. McGrath Joseph Edward. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.
- [56]. Morgan David L. 1996. Focus groups. *Annual review of sociology* 22, 1 (1996), 129–152.
- [57]. Murata M, Ahmetovic D, Sato D, Takagi H, Kitani KM, and Asakawa C. 2018. Smartphone-based indoor localization for blind navigation across building complexes. In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom). 1–10.
- [58]. Nguyen Brian J, Kim Yeji, Park Kathryn, Chen Allison J, Chen Scarlett, Van Fossan Donald, and Chao Daniel L. 2018. Improvement in patient-reported quality of life outcomes in severely visually impaired individuals using the Aira assistive technology system. *Translational Vision Science & Technology* 7, 5 (2018), 30–30.
- [59]. Olson Gary M and Olson Judith S. 2000. Distance matters. *Human-computer interaction* 15, 2–3 (2000), 139–178.
- [60]. Olsson Thomas, Huhtamäki Jukka, and Kärkkäinen Hannu. 2020. Directions for professional social matching systems. *Commun. ACM* 63, 2 (2020), 60–69.
- [61]. Petrie Helen, Johnson Valerie, Strothotte Thomas, Raab Andreas, Michel Rainer, Reichert Lars, and Schalt Axel. 1997. MoBIC: An aid to increase the independent mobility of blind travellers. *British Journal of Visual Impairment* 15, 2 (1997), 63–66.
- [62]. Ponchillia Paul E and Ponchillia Susan Kay Vlahas. 1996. Foundations of rehabilitation teaching with persons who are blind or visually impaired. American Foundation for the Blind.
- [63]. Ranjan Abhishek, Birnholtz Jeremy P, and Balakrishnan Ravin. 2006. An exploratory analysis of partner action and camera control in a video-mediated collaborative task. In Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. 403–412.

- [64]. Real Santiago and Araujo Alvaro. 2019. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors (Basel, Switzerland)* 19, 15 (02 Aug 2019), 3404. 10.3390/s1915340431382536. [PubMed: 31382536]
- [65]. Rentzsch Robert and Staneva Mila. 2020. Skills-Matching and Skills Intelligence through curated and data-driven ontologies. In *Proceedings of the DELFI Workshops*.
- [66]. Rocha Sebastião and Lopes Arminda. 2020. Navigation based application with augmented reality and accessibility. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. 10.1145/3334480.3383004
- [67]. Rosenblat Alex and Stark Luke. 2016. Algorithmic labor and information asymmetries: A case study of Uber’s drivers. *International journal of communication* 10 (2016), 27.
- [68]. Sato Daisuke, Oh Uran, Naito Kakuya, Takagi Hironobu, Kitani Kris, and Asakawa Chieko. 2017. NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 270–279.
- [69]. Scheggi Stefano, Talarico A, and Prattichizzo Domenico. 2014. A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. In *22nd Mediterranean Conference on Control and Automation*. IEEE, 20–23.
- [70]. Shaw Marvin E. 1981. *Group dynamics: The psychology of small group behavior*. McGraw-Hill College.
- [71]. Simons Rachel N, Gurari Danna, and Fleischmann Kenneth R. 2020. “ I Hope This Is Helpful” Understanding Crowdworkers’ Challenges and Motivations for an Image Description Task. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26.
- [72]. Microsoft Soundscape. 2020. A map delivered in 3D sound. <https://www.microsoft.com/en-us/research/product/soundscape/>.
- [73]. Steiner Ivan Dale. 1972. *Group process and productivity*. Academic press New York.
- [74]. Stephan Frederick F. 1952. The relative rate of communication between members of small groups. *American Sociological Review* 17, 4 (1952), 482–486.
- [75]. Stephan Frederick F and Mishler Elliot G. 1952. The distribution of participation in small groups: An exponential approximation. *American sociological review* 17, 5 (1952), 598–608.
- [76]. Tang John C. 1991. Findings from observational studies of collaborative work. *International Journal of Man-machine studies* 34, 2 (1991), 143–160.
- [77]. Tversky Barbara. 1993. Cognitive maps, cognitive collages, and spatial mental models. In *Spatial Information Theory A Theoretical Basis for GIS*, Frank Andrew U. and Campari Irene (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 14–24.
- [78]. Wiener WR, Welsh RL, and Blasch BB. 2010. *Foundations of Orientation and Mobility*. Number v. 1 in *Foundations of orientation and mobility*. AFB Press. <https://books.google.com/books?id=nYENqA5LZKUC>
- [79]. Xie Jingyi, Reddie Madison, Lee Sooyeon, Billah Syed Masum, Zhou Zihan, Tsai Chun-hua, and Carroll John M. 2022. Iterative Design and Prototyping of Computer Vision Mediated Remote Sighted Assistance. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–40.
- [80]. Xie Jingyi, Yu Rui, Lee Sooyeon, Lyu Yao, Billah Syed Masum, and Carroll John M. 2022. Helping Helpers: Supporting Volunteers in Remote Sighted Assistance with Augmented Reality Maps. In *Designing Interactive Systems Conference*. 881–897.
- [81]. Yoon Chris, Louie Ryan, Ryan Jeremy, Vu MinhKhang, Bang Hyegi, Derksen William, and Ruvolo Paul. 2019. Leveraging augmented reality to create apps for people with visual disabilities: A case study in indoor navigation. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 210–221.

CCS CONCEPTS

- **Human-centered computing** → *Accessibility; Empirical studies in accessibility.*

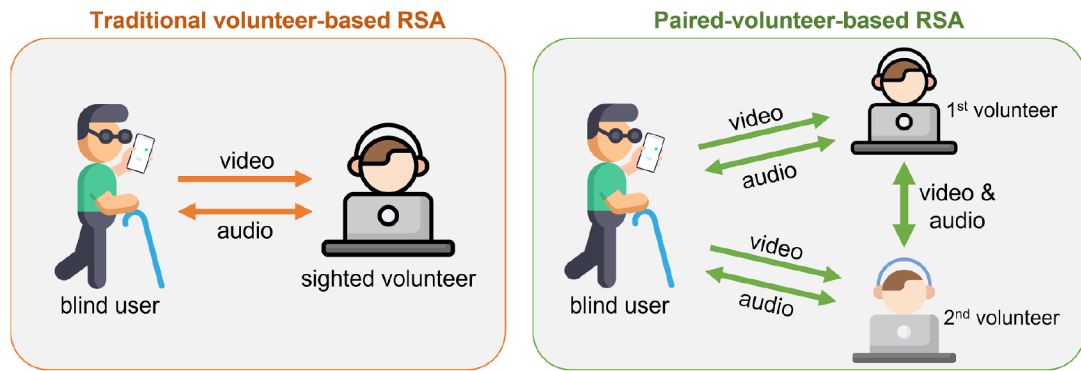


Figure 1:

A comparative illustration showcasing the distinction between traditional remote sighted assistance (left) – where a single sighted volunteer aids a blind user – and our proposed approach of paired-volunteer remote sighted assistance (right), involving two sighted volunteers working together to provide collaborative support to the same user.

Task categories, a sampler of frequently-requested tasks under each category in current RSA, and a sampler of envisioned tasks under each category in paired-RSA.

Table 1:

Task Category	Commonly Rendered Tasks in RSA	Envisioned Tasks for Paired-RSA
C1: Reading printed text [12]	Reading labels of bottles and cans, postal mails, a children’s book, a greeting card, a restaurant menu	Reading multilingual text
C2: Describing visual media	Explaining colors, memes, exercise videos	Appreciating artwork; interpreting aesthetics videos; describing jewelry
C3: Finding lost or dropped objects [12]	Picking up keys, wallets, and wedding rings	Find the dropped headphone
C4: Kitchen assistance [50]	Distinguishing between food items; checking the freshness of products	Cooking assistance
C5: Learning or setting up appliances [50]	Navigating TV menus, using (3D) printers, operating vending machines	Expert assistance for doing homework; learning how to operate household appliances and computers
C6: Shopping [12, 51, 79]	Learning the aisles of a grocery store, selecting food with similar packages	Shopping grocery or clothing in real-time; online shopping
C7: Navigational assistance [44, 50, 51, 80]	Looking for landmarks, identifying road conditions, finding and verifying rideshares	High-risk and wayfinding tasks
C8: Assistance in pursuing hobbies and leisure activities		Playing video games; making origami; or paper art
C9: Style and beauty assistance	Checking makeup, identifying outfits	Applying makeup, matching outfits

Blind participants' demographics. They were the users (U1 to U9) in our proposed paired-RSA.

Table 2:

ID	Gender	Age Group	Condition of Vision Impairment	Age of Onset	Occupation Type
U1	F	35–40	Retinopathy of prematurity, Glaucoma, and cataracts	Legally blind since birth; completely blind 8 yrs ago	Unemployed
U2	M	20–25	Central vision affected, color perception	14 yrs old	Student
U3	F	40–45	Retinitis pigmentosa	22 yrs old	Vision rehabilitation therapist
U4	M	40–50	Retinopathy prematurity	At birth	IT professional
U5	F	35–40	Cone-rod dystrophy, probably 20/600 acuity	At birth	Program director at a nonprofit org
U6	F	20–25	Microphthalmia	At birth	Student
U7	M	25–30	Completely blind, optic nerve atrophy, leukemia	At birth	Sales specialist
U8	M	30–35	Glaucoma	9 yrs old	Unemployed
U9	F	25–30	Completely blind, microphthalmos, vascularized corneas, and cataracts.	At birth	Student

Sighted volunteers' demographics. They were the volunteers (V1 to V8) in our proposed paired-RSA.

Table 3:

ID	Gender	Age Group	Number of Calls Received	Time as a Sighted Volunteer	Occupation Type
V1	M	45–50	50 BME calls	4 yrs, 2500 hrs as an Aira agent	Administrative assistant
V2	F	25–30	3	3 yrs as a BME volunteer	Healthcare
V3	F	20–25	1	6 mos as a BME volunteer	Student
V4	F	25–30	At least 10	2 yrs as a BME volunteer	In-between jobs
V5	F	20–25	More than 12	2 yrs as a BME volunteer	Student
V6	M	50–55	More than 10	3 yrs as a BME volunteer	IT
V7	F	30–35	More than 5	2 yrs as a BME volunteer	Teacher
V8	F	35–40	More than 300	4 yrs as an Aira agent	Clinical medicine

Table 4:

Sessions and tasks information. Note that there is no task under category C4 (kitchen assistance), as no user asked for it.

Session	Participants	Tasks and Task Category (C1 to C9)
1	U1, V2, V3	Read the multilingual text (Korean, Spanish, and Japanese) (C1); Appreciate Van Gogh's paintings (C2).
2	U2, V4, V5	Describe music videos (C2); Find dropped headphones (C3); Read street signs (C7); Make a paper airplane (C8).
3	U3, V6, V7	Describe video tutorials of bead artwork (C2); Describe color choices for online yarn shopping (C6); Describe patterns of bead artwork (C8).
4	U4, V1, V3	Learn an electric grill's control panel (C5); Make a paper dog and a paper bird (C8).
5	U5, V2, V7	Describe patterns of jewelry (C2); Help with makeup (C9); Matching outfits (C9).
6	U6, V1, V4	Real-time grocery shopping (C6).
7	U7, V2, V6	Describe video tutorials of a paper butterfly (C2); Amazon shopping (C6).
8	U8, V4, V7	Label microwave buttons (C5); Sort vinyl records (C8).
9	U9, V3, V5	Read mail (C1); Label temperatures on toaster and oven (C5).
10	U2, V1, V8	Describe photography (C2); Identify item locations in a local store (C6); Navigate around campus and learn locations of classrooms (C7).
11	U7, V5, V6	Read customer reviews (C1); Amazon shopping (C6).

Five dimensions (D1 to D5) in our taxonomy of tasks completed in the study. Note that there is no task under category C4 (kitchen assistance).

Table 5:

Task	Task Examples and Category (C1 to C9)	Type (D1)	Shared Content (D2)	Temporal and Cognitive Demand (D3)	Special Skills Required (D4)	Division of Labor (D5)	Suitableness for Paired-RSA
Reading	Reading mail, multilingual text (C1)	Objective	Screen or live camera feed	Low	Yes	Serial	Conditional: Yes, if special skills needed (e.g., technical knowledge, multilingual); No, otherwise
Describing Visual Media (aesthetics)	Appreciating artwork (C2)	Subjective	Screen	Low	Yes	Serial	Yes, as it offers multiple perspectives
Describing Visual Media (videos)	Describing music videos, DIY videos (C2)	Subjective	Screen	High	No	Serial	Conditional: Yes, if blind participants can pause the videos; No, otherwise
Finding Dropped Objects	Finding a dropped headphone (C3)	Objective	Live camera feed	Low	No	Parallel	No
Help with Household Appliances	Labeling microwaves, learning functions of a new grill (C5)	Objective	Live camera feed	Low	Yes	Hybrid	Yes, brainstorming solutions
Grocery Shopping	Purchasing instant noodles (C6)	Objective, Subjective	Live camera feed	High	No	Parallel	Conditional: Yes, with a silent partner; No, otherwise
Online Shopping	Shopping headphones on Amazon (C6)	Objective, Subjective	Screen	Low	No	Hybrid	Yes, more opinions
Navigation	Walking around campus (C7)	Objective, Subjective	Live camera feed	High	Yes	Parallel	Conditional: Yes, with a silent partner; No, otherwise
Crafting	Making origami (C8)	Subjective	Live camera feed	Low	No	Serial	Yes, clarification for blind participants
Fashion Help	Applying makeup, matching outfits (C9)	Subjective	Live camera feed	Low	Yes	Serial	Yes, more opinions