Interdependence in Action: People with Visual Impairments and their Guides Co-constituting Common Spaces

BEATRICE VINCENZI, Centre for HCI Design, City, University of London, UK
ALEX S. TAYLOR, Centre for HCI Design, City, University of London, UK
SIMONE STUMPF, Centre for HCI Design, City, University of London, UK

Prior work on AI-enabled assistive technology (AT) for people with visual impairments (VI) has treated navigation largely as an independent activity. Consequently, much effort has focused on providing individual users with wayfinding details about the environment, including information on distances, proximity, obstacles, and landmarks. However, independence is also achieved by people with VI through interacting with others, such as in collaboration with sighted guides. Drawing on the concept of *interdependence*, this research presents a systematic analysis of sighted guiding partnerships. Using interaction analysis as our primary mode of data analysis, we conducted an empirical, qualitative study with 4 couples, each made up of person with a vision impairment and their sighted guide. Our results show how pairs used interactional resources such as turn-taking and body movements to both co-constitute a common space for navigation, and repair moments of rupture to this space. This work is used to present an exemplary case of interdependence and draws out implications for designing AI-enabled AT that shifts the emphasis away from independent navigation, and towards the carefully coordinated actions between people navigating together.

CCS Concepts: • **Human-centered computing** → **Accessibility**; *Accessibility technologies*; *Accessibility design and evaluation methods*; *Empirical studies in accessibility*;

Additional Key Words and Phrases: Assistive technology; AI; interaction analysis; interdependence; vision impairment; sighted guiding.

ACM Reference Format:
Beatrice Vincenzi, Alex S. Taylor, and Simone Stumpf. 2021. Interdependence in Action: People with Visual Impairments and their Guides Co-constituting Common Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 69 (April 2021), 33 pages. https://doi.org/10.1145/3449143

1 INTRODUCTION

People living with varying forms of visual impairment (VI) represent a significant population worldwide [47]. Assistive Technology (AT) has become a common part of daily life for this population, supporting them in activities and routine tasks, such as reading text, navigating the web, using smartphones, and so on [12, 24, 41, 46, 70]. More recently, artificial intelligence (AI) has been promoted as a means for extending these ATs. For example, increased attention has been given to designing and developing AI-enabled AT to aid independent navigation for people with visual impairments. In this research, much effort has been dedicated to providing the individual with

Authors' addresses: Beatrice Vincenzi, beatrice.vincenzi@city.ac.uk, Centre for HCI Design, City, University of London, UK, London, UK; Alex S. Taylor, alex.taylor@city.ac.uk, Centre for HCI Design, City, University of London, UK, London, UK; Simone Stumpf, simone.stumpf.1@city.ac.uk, Centre for HCI Design, City, University of London, UK, London, UK.

2573-0142/2021/4-ART69 $15.00
https://doi.org/10.1145/3449143

wayfinding guidance and complimentary information about the physical features in an environment such as one's proximity to obstacles, curbs, hazards and landmarks [3, 22, 23, 29, 39, 42].

In the following, we aim to engage with the research operating at the intersections of AI and its use in assistive guidance technologies. However, in contrast to much of the prior work, our research is informed by a perspective that stresses the *interdependencies* between people with vision impairments and others in their surroundings, what might be thought of as the *collaborative work* in CSCW terms. Our thinking draws on scholarship in disability studies that highlights the significance of ***inter**dependence* as opposed to independence [7, 18, 32]. It also builds, in particular, on the notion of an *interdependence frame* as proposed by Cynthia Bennett, Erin Brady and Stacy Branham [5]. These ideas from disability studies and from Bennett et al. emphasise the collective work done by people with disabilities and others to achieve access and independence. Further, they make clear these relationships are not only between people, but also include the interplay with AT and the environment. It is through this line of theorising and reflection that the presented work aims to rethink the role of AI in AT, considering, in particular, how exactly AI might become a valuable resource in the unfolding relations between people, technology and their surroundings.

Taking up the call from Bennett et al. [5], we consider how the sighted guide relationship offers an example of, one, people with VI working with others—in this case their sighted guides—and, two, an interdependent partnership potentially open to AI-based interventions. As it is routinely described, the sighted guide technique involves a sighted person guiding using touch and voice to forewarn a person with a visual impairment about kerbs, steps, obstacles, etc. The guidance in this sighted-guide configuration is somewhat standardised: if someone needs support, the guide bends the arm parallel to the ground and offers their arm or elbow to the person being guided. By holding the arm's guide and having the guide one-half step ahead, the companions can navigate together [52]. A primary concern in our research has been with how, in practice, such collaborative work is accomplished. We ask how guides and people with vision impairments work together to competently navigate their surroundings—how interdependence is achieved *in action*? Rather than presuming an instrumentalised idea of independent navigation (i.e. that navigation is a problem AI can solve), we aim to consider what role AI might play in augmenting and extending the interdependencies. Thus, we ask not how we might "solve" the "problem" of navigation, but how people's collective capacities–their capacities for working together–might be enhanced [6].

In response to this thread of thinking, we conducted an empirical, qualitative study to explore how interdependence is interwoven into the sighted guide relationship. We recruited four pairs of participants composed of a person with a visual impairment and their guide to take daily routine journeys using the sighted guide configuration. Wearing two body cameras, participants video recorded their interactions in the settings they navigated.

As discussed in the Findings and Discussion sections, the results of this study provide lessons for augmenting the sighted guide relationship and reorienting the design and development of AI-enabled assistive technology. Broadly, our work points away from efforts to supplement simple notions of individual autonomy and agency, and instead sees these as achieved in and through the coordination between actors and their surroundings. As such, autonomy and agency are approached not as static attributes to be bestowed on a person with a visual impairment by an AI-enabled AT. Rather than seeking to mimic the role of the guide (and, in effect, replace the guide), we show how AI can play an important role in extending the capacities both parties use and acquire *together*.

Our work makes four main contributions: (1) we show how sighted guiding is a collaborative venture, where people use a variety of interactional resources to establish a common space that they can navigate through; (2) we provide a rich and detailed understanding of how these interactions constitute interdependence, in which agency moves fluidly between people; (3) we show evidence of ruptures in these interdependent relations and how they are repaired through the use of interactional resources; and (4) we provide implications for designing AI-enabled assistive guidance technology, placing the emphasis on further enabling collaboration rather than simple and reductive ideas of autonomy and agency.

2 RELATED WORK

We now present the state-of-the-art in AI-based assistive technology aimed at navigation, and how the *interdependence* frame [5] can be used as a way to think differently about designing assistive technology, especially when applied to the sighted guide relationship.

2.1 AI-based Technology Used to Assist Independent Navigation

Assistive technology aimed at people with visual impairments seeks to support them by improving access to technology, their independence and, in turn, quality of life. Prior work has designed solutions addressing a wide range of activities such as reading, writing, gaming and navigation. Currently, these AT solutions work by augmenting or replacing vision [67, 70], for example, applications that help people with limited vision to magnify any text, image or video that the device captures [46, 70] or transforming graphics, images, or text into audio and synthesized speech through screen readers or haptic information to braille devices [2, 49, 65]. In addition, there is also a drive to leverage artificial intelligence (AI) for some tasks that rely on visual information, such as the identification of objects, barcodes and currency [12, 24, 41].

Independent navigation for people with visual impairments is considered a major challenge, drawing significant attention from the research communities in both HCI and AI [23]. Here, AI is often used to solve a functional task, where the user follows instructions to successfully reach a destination. Hence, research has focused on supporting how users navigate physical spaces, and aiding in the identification and proximity of walls, curbs, obstacles, streets, etc. often using beacons [3, 34, 39, 63, 71] or computer vision systems [3, 29, 34, 66]. For example, the *Cities Unlocked* project [42]—a collaboration with *Microsoft* and *Guide Dogs for the Blind*—proposes a wearable headset connected to a smartphone application to receive information about a current location, surrounding streets and landmarks, and help to explore the surrounding environment and reach a destination. The *NavCog* system is exclusively for indoor use and relies on Bluetooth low energy (BLE) beacons, installed in the environment, to estimate the user's position and provide turn by turn instructions [22]. A recent application, *BBeep*, predicts the future position of pedestrians and tracks the user's path; when it predicts a potential collision between the traveller and the pedestrian, the system alerts both the user and the nearby pedestrian [31].

Again, these solutions target autonomous travel, treating it as a functional task. As a consequence they place the emphasis on more accurate information about the environment and wayfinding to provide greater independence to the user. Relationships with other people (e.g., pedestrians, assistants, guides, friends, etc.) are given little consideration, with others' bodies largely treated as physical masses, either moving or stationary, and as something to be avoided.

2.2 A different Perspective: the Interdependence Framework

Recent research has stressed the agency of people with disabilities in their collaborative work with others, and demonstrates how both access and independence are achieved through this interdependence [5, 6]. Thus, assistive technologies should not be approached as a "*gap between disabled bodies and environments designed for non-disabled people*" [5, p.161], but as an aspect of the ongoing interplay between different actors and the specificities of any one setting. *Be My Eyes* [17] and *BeSpecular* [36] offer two compelling examples of services that open up a space for such mutual and collaborative work, providing people with visual impairments remote access to crowd sourced communities of sighted users.

The interdependence concept draws from and has been intensely discussed in disability studies. What might be captured loosely as the *relational perspective* [7, 32] understands disability as being in continual production, where bodies, technologies, settings, etc. are unceasingly entwined to make actors more or less able. Moser [45] shows how disability is not something fixed within the body, but is manifested through interactions with the environment, other people and technology: "*disability is not something a person is, but something a person becomes*". Similarly, Goodwin [20] presents systematic analysis of a person with aphasia and who is able to speak only three words but nonetheless acts as a competent speaker. Through Goodwin's work we see how a complex conversation can unfold with few words, but in combination with body movements and gestures, and the interplay with the talk and actions of other actors. Ability and disability are then capacities made possible through the relations with others.

While previous HCI research has paid attention to the social features of settings, it has primarily been related to the ability-based design of technology or the social implications of existing ATs. For instance, previous perspectives highlight mismatches between the different ways sighted and non-sighted people interact with their surroundings, and thus how appropriate feedback is important in navigation [69]. The form and function of an AT has also been considered in terms of how it influences social interactions, self-perception and social acceptability [59, 60]. In light of the research in disability studies, such approaches present relatively static versions of the relationships between the actors, and what capacities they may have to work together.

Relationality and the interdependence perspective thus opens up opportunities to think differently about the design of AT and consequently about the use of AI in this context. Bennet at al.'s frame [5], in particular, sees AT as a further way to extend the relations between one another, focusing on how actors are made more or less able, relationally, through other actors and with/through AT. Other recent work further illustrates the complexities of these social relations and how (dis)ability disappears or emerges through them. Thieme et al. [64], for example, examine how people with visual impairments negotiate their abilities and how they make sense of the environment through different resources and collaboration with others in several contexts during the Rio Paralympics. Similarly, research has shown the collaborative work done involved between people with VI and people who are sighted to co-create an accessible home environment [8], and how a mixture of abilities can operate together to achieve tasks [6, 9]. However, technologies and applications which go beyond the accomplishment of individual tasks, and pay particular attention to social activities and relations, remain unexplored in the design of AT.

The interdependence frame would seem especially applicable in understanding sighted guiding. To approach this aided guidance from a perspective of autonomous travel would be to reduce the "problem" of navigation to a sequence of steps and the movement from one place to another, and ultimately, to look to solutions that replace the sighted guide with an AI system. A compelling paper from Due and Lange [15] offers some relevant lessons from an ethnographic and ethnomethodologically orientated study of people with vision impairments navigating with guide dogs. The work presents an analysis of visually impaired people competently navigating their environments with guide dogs. Specifically, it highlights a distinction between the use of a white cane, seen as "problem-oriented", and the guide dog, seen as "solution-oriented". So, in contrast to the white cane, guide dogs do not inform their companions about obstacles but instead help to establish safe routes. This draws particular attention to the interdependent relations between the person and their guide dog.

Similarly, when viewed in terms of interdependence, sighted guiding represents a clear example of people working together and, thanks to their collaboration, being able to successfully coordinate their actions to move through space. To think with interdependence is then to recognise movement and space not in strictly euclidean, geometric terms but as something that is co-produced and mutually orientated to to accomplish activities like navigation [14]. Moreover, sighted guiding provides an enjoyable way to explore familiar and unfamiliar environments across the visually impaired community [61], for building and intensifying relationships [18, 32], and for feelings of freedom [37]. It is thus through sighted guiding, as an illustrative case, that we aim to develop and refine the evolving work into interdependence and approach AI's potential in AT from a different perspective. Broadly, our aim is to develop (1) insights into how people with VI and their sighted guides navigate together successfully, and how this experience exemplifies interdependence; and (2) identify implications for designing AI-based assistive technology to extend the interpersonal interactions which aligns to the interdependence framework.

3 STUDY METHODS

To investigate the current sighted guiding relationship, we conducted an empirical study in which we invited people with visual impairments and their sighted guides (with whom they usually travel) to video-record their real-world journeys using body-worn video cameras. As we assumed interpersonal coordination to be a key element in sighted guiding, we analysed our data using interaction analysis to examine the details of how guiding and being guided were accomplished. We now provide details on the participants in our study, how the study was structured and carried out, and how we collected and analysed the data.

3.1 Participants: People with Visual Impairments and their Sighted Guides

We recruited pairs composed of adults: a person registered severely visually impaired (blind) or visually impaired (partially sighted) and a sighted guide. Both had to know each other through guiding for at least 3 months. This allowed us to ensure that they had some experience of guiding each other, to investigate their established interactions as safely as possible, and also to observe their relationship with each other that might differ between pairs in terms of harmony, care and mutual understanding.

Participants were recruited through adverts via social media, emails to existing contacts and printed flyers. Since we were targeting both sighted and visually impaired people, different formats were essential; we made considerations for the diversity of vision impairments and accessibility on all electronic materials. We excluded people with cognitive or mobility impairments from our study that could have prevented them giving informed consent or being able to travel outside the home without additional assistive mobility aids. Accessible participant information sheets and consent forms were emailed to participants in advance. Informed consent was obtained at the initial face-to-face meeting between each pair and the researcher. Approval for this study was granted by the Computer Science Research Ethics Committee at City, University of London. In appreciation of participants' contribution to the research we offered a £25 voucher per person.

Four participant pairs took part in the study and Table 1 reports a summary of demographic information we gathered. The names used throughout the article are pseudonyms. As shown in Table 1, these pairs knew each other between 2 and 20 years, ranging from only knowing each other through guiding to being married. For instance, *Alan* and *Nick* have a mentor-friendship and they have known each other for one to two years. Nick is the main founder of a charity and Alan is involved in a creativity project run by Nick's charity. Nick guides Alan every week from the tube station to Nick's studio and *vice-á-versa.* Nick said that guiding Alan every week is only a small thing that happens during the journey, there is much more going on and travelling with Alan is a great occasion to build their friendship. In contrast *Megan* and *Jack*'s relationship is quite different. They have known each other for 2 years and Jack has been guiding Megan for 1 year from time to time, but only when Megan needs particular assistance for travelling in unfamiliar and crowded places. Megan is quite independent and ordinarily uses a long cane in her daily journeys. While some of the participants had no sight, some participants were able to distinguish light and dark. In addition to sighted guiding, they also used tools such as white canes and guide dogs to travel on their own. All participants living with sight loss had an active life, using assistive technology such as screen readers, voice synthesizers, etc.

Pair *Person with VI	Relationship	Info about their vision	Aid for travelling	Assistive Technology
Megan* and Jack	2 yr - Guiding relationship	R: no vision L: blurry vision since birth	Long cane and sighted guide	Screen reader Voice synthesizer Magnifier
Alan* and Nick	2 yr - Mentorship	No vision since birth	Long cane and sighted guide	Screen reader Voice synthesizer Braille
Luke* and Alice	7 yr - Close friends	Light and dark for 27 yr	Long cane and sighted guide	Screen reader Voice synthesizer Apps on mobile
George* and Sara	20 yr - Married couple	R: only central vision L: blurry-no color for 35 yr	Guide dog and sighted guide	Screen reader Voice synthesizer Large monitor High contrast on computer/mobile

Table 1. Demographic Information.

3.2 Study Setup

The study took place over a week, consisting of an initial meeting and the pairs recording journeys (Figure 1).

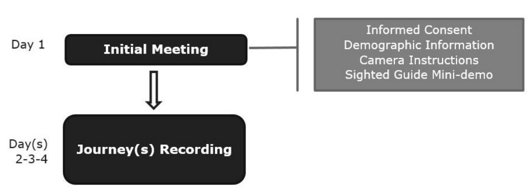


Fig. 1. Research study procedure.

During the initial meeting between the pair and the researcher, we obtained informed consent and gathered demographic information. We then gave instructions on how to wear and use the cameras. Video recordings were captured using two body cameras, as illustrated in Figure 2. Cameras were worn by both participants using a harness at chest or shoulder height. The right position and orientation of the camera depended mainly on the difference in height between participants and how close they usually walk. We were interested in capturing the physical connection between the guide's arm and the visually impaired person's arm and the upper back of the guide. These parts of the body were captured by the visually impaired person's body camera (see Figure 2a). The second body camera, worn by the sighted guide, was aimed to record the guide's perspective. We also did a mini-demo to verify that the camera configuration worked in relation to how pairs travelled together, using a wifi connection between the camera and the researcher's phone app.

To minimise the risk of unauthorised access to the video data, we used two *S-EYE* body cameras from *Shelleyes Group*, with built-in encryption protocols. These cameras were also highly appreciated by participants because they were accessible, having a simple layout, big buttons with different textures, and a variety of audio and vibration feedback features. The pair then had three days to record one or more journeys, with each journey lasting at least 20 minutes to ensure enough data. The journeys were chosen by the pairs themselves; the only stipulation was that the journeys be familiar to them, for example, going grocery shopping, to a museum, to a GP appointment, a coffee to meet friends, etc. This guidance was given because we were interested in the routines of everyday life and in capturing the ordinary ways people manage their partnerships. In addition this choice allowed us to mitigate ethical concerns related to participants' safety.

(a) VI person wearing camera.

(b) Sighted guide set up.

(c) Frame from video data showing both different perspectives: VI person perspective on the left and sighted guide on the right.

Fig. 2. Body cameras set up.

3.3 Data Collection and Analysis

Across four participants pairs, we collected 6 journey records in 23 shorter clips of 10 minutes (i.e., a total of 4 hours of journey data). We then pre-processed all the videos and audio so that the guide's video and the guided person's video were displayed, visually, side-by-side and synchronised (see Figure 2c).

The data was analysed with interaction analysis, which is a qualitative research method, well established in studies of the workplace [26, 27, 40]. The theoretical underpinnings of this approach, owing much to conversation analysis and ethnomethodology, offer a means of understanding how interaction is made intelligible and thus consequential to the members of a setting, so, for example, how people cross a road at traffic junctions without constant collisions [35] or musicians achieve synchrony [68]. Over the past two decades it has also emerged as a research method to investigate specialized forms of social activities [26] and more recently in everyday and technology-mediated settings [11, 33, 50].

Interaction analysis is also adopted in HCI and CSCW research as a method of analysis to inform technology design, for example, to describe how tourists work together in groups, collaborate around maps and guidebooks [10], how passers-by interact with urban technology in public space [4], and how people collaborate and jointly interact with other mobile technologies while driving [48]. Especially relevant to the work we present, Due et al. [16] show how people are able to detect obstacles during navigation using a white cane and describe design implications for future technology. More generally, interaction analysis has been used to study disability in context. For example, Goodwin [20] (a central contributor to interaction analysis and conversation analysis), has used the research method to demonstrate that limitations to vocalised speech can in many ways be overcome through pointing gestures, head turning, and gaze in collaboration with others in a setting.

In practice, interaction analysis relies on repeated and careful re-watching of recorded video to produce detailed transcripts of spoken and interpersonal interactions, including non-verbal communication (for example how people orient their bodies, which gestures they use, what they are pointing to and so on) to investigate "*the ways in which specialised tasks and activities are accomplished through embodied activity, activities that involve the interplay of talk, visible conduct and the use of various objects and artefacts, tools and technologies*" [25].

In our research, because of the depth and detail of interaction analysis, we chose to focus our attention on salient segments/excerpts from across the 23 clips we collected. A segment or excerpt is a short video from our data, which lasts on average 20 seconds. Segments were chosen through repeatedly watching the recorded data, individually and as a team, and identifying parts we felt presented compelling examples of coordinated interaction. In particular, we looked for examples of how our participants worked together to move through/past obstacles such as narrow gaps, curbs, cars, and other people, and how they managed barriers, or crossing thresholds such as moving in/out of buses and stores. We also paid particular attention to moments which emphasised the work done by pairs to co-locate themselves in the guiding configuration while for example shopping, paying at the counter or beginning/ending a new journey. A total of 40 segments were analysed in detail in this way. In a similar way to how interaction analysis findings are reported in CSCW [11, 28] and in other research communities [20, 33], for this article we chose to present segments that best illustrate the themes documented in Section 4. Specifically, segments were selected that featured different aspects of the use of multimodal resources to co-constitute a common space; the interdependent work done in sighted guiding; and instances of ruptures and repairs to common space.

4 FINDINGS

Below, we select a set of 6 relevant segments/excerpts out of 40 to present our findings. For each segment, we present a short description to introduce the participant pair and the context of their journey; this is followed by a detailed analysis. A transcription of the conversation, along with images to show non-verbal communication, is attached at the end of each sub-section. Appendix A provides a detailed explanation of symbols used in the transcription.

Through the following analysis we aim to develop three ways of making sense of the sighted guide relationship paying particular attention to the ways interdependence plays into their coordinated actions and movements. Broadly, we show (i) how a common space is co-constituted between people with vision impairments and their guides; (ii) how, as a form of interdependence, this work together is interwoven into the ordinary and unfolding sequence of interactions; and (iii) how the unfolding relations in/through space are subject to rupture and open to repair.

4.1 Co-constituting a Common Space

In the first of our analysed segments, we describe how pairs work together to establish a common space to move in and through. Specifically, we show how a mutual orientation to talk, body movements and gestures, and other objects help to constitute a space that can be navigated together. Noteworthy will be the coordinated actions between the pairs, what we wish to highlight as the *co-constituting of space*. It is through these coordinated orientations and actions that we will show *how*, exactly, interdependent relations are accomplished between people with vision impairments and their guides.

4.1.1 "Come on Step Down": Through this first excerpt (see Segment 1), we begin our analysis by considering how both talk and objects—objects like canes—play into constituting a common space between pairs. Luke (visually impaired person - VIP) and Alice (Guide - G) are walking in their neighbourhood as the conversation turns to the nice weather over the last few days. The friends walk side by side—Luke is holding Alice's arm with his right hand and holds a long white cane on his free (left) arm. As they approach a sidewalk curb (line 12), Alice announces *"come on ↑step down Luke"*. This utterance and Alice's *"curb and bicycle ↑step up"* in a subsequent turn (line 17), briefly interrupt the ongoing conversation.

Notice, first, how talk itself is being used to coordinate actions and establish a common understanding of space. Well-established works in conversation analysis show that *how* talk is conducted can serve as a resource in the organisation of turn-taking [54]. For example, how interlocutors say what they say can indicate the way a sequence of turns occurs, the opening up and closing of topics of talk [58], and the repair of troubles in talk [57]. (In conversation analysis, repair refers to the ways in which talk is kept on track. Repair techniques can be displayed through a subsequent spoken turn, but also using a variety of non-lexical speech perturbations, cut-offs, sound stretches, onomatopoeic words, etc.). What these works allow us to identify are the relevant methods used to establish the change in a conversation's topic and how some changes are achieved turn-by-turn, coordinated through a mutual exchange of verbal and non-verbal signals.

In the segment described above, Alice forewarns Luke of the sidewalk's curb through a change in tone and raising pitch in talk (i.e. *"↑step"*). This adjustment in talk provides Luke with a cue, indicating the utterance refers not to the ongoing topic—the weather—but another matter, in this case to do with navigation. Further, phrases such as *"come on"* and *"Luke"* are used to emphasise a discontinuity between the primary topic and the navigational cues she provides. This, we might suggest is one reason why Alice does not have to begin a lengthy explanation of the approaching curb, but is simply able to say *"step up"* or *"step down"* to establish a shared sense of space.

Of course, the act of walking together, and the mutual awareness that obstacles like curbs and steps must be managed together, also attributes phrases like *"step up/down"* with an *indexical* quality. That is, we see that *when* something is said can serve as an *index* to a feature in the environment without an explicit need to describe it in full (similar to saying, for example, *"that"* and pointing at something). The critical point here is that a space for Luke and Alice is being continually composed or established through an ongoing and interwoven set of mutually produced and intelligible resources.

Let us consider one further point from this excerpt. Above, it seems Luke's white cane plays a consequential role. The cane has just touched the pavement (Fig. 3d) when Alice says: *"step down Luke"* and, again, is swept against the sidewalk corner when she announces *"step up"* (Fig. 3f). The importance of the white cane as a resource to detect troubles and obstacles has been explored by Due et al. [15, 16]. In their work, they observe how visually impaired people use their canes to skilfully and competently navigate while traversing known routes; this illustrates the variety of resources being brought to bear on navigation. What the interchange between Luke and Alice adds to this previous research is a recognition of the coordinated actions between a pair walking together: the synchronisation in time and space—between Alice's talk, Luke's cane sweeping, and their shared walking pace—looks to be critical to their successful navigation and for Luke to confidently take the following next step up. Notably, there is no hesitation on Luke's part, and there are no pauses or the need for other repairs [54] after Alice's alert. The cane becomes relevant because what Luke perceives through it is combined with Alice's utterances, her change in pitch and tone, and their common movement. Again, altogether, we witness a rich and multi-threaded composition of space, and a space constituted together.

Segment 1 - **Come on step down** (Luke and Alice)

```
1        L:   that's nice, isn't it?
2        A:   that felt like the other week
3        L:   ye::ah
              0.4
              Fig 3a
              ▼
4        L:   this will (of an autumn) not just straight into winter
5        A:   yes::
6        L:   really nice
7        A:   I know
8        A:   o::h () we are going to (X) today Luke // too much (checking) traffic probably
9        L:   *(what's the plan)*
              Fig 3b
              ▼
10       L:   still traffic // alright
11       A:   *((coughing))*
                                                   Fig 3c
                                                   ▼
12       L:   in theory it may be a good go for the camera=
                                      Fig 3d
                                      ▼
13  ⟶   A:   =come on ↑step down Luke=
14       L:   =( ) this is illegal I mean // ehm
15       A:   *((laughing))* it is not illegal Luke it is perfectly safe
16       L:   it is not true // ( )
                               Fig 3e    Fig 3f
                               ▼         ▼
17  ⟶   A:   *curb and bicycle, ↑step up*
              0.5
18       A:   ye:::ah we made it
19       L:   yap
20       A:   so far
21       L:   one piece
```

(a) L and A walking on the pavement

(b) L and A approaching a curb

(c) step down L

(d) The cane is touching the pavement

(e) A alerting L about the curb

(f) step up

Fig. 3. Segment 1 - Come on step down (Luke and Alice).

4.1.2 *"Go Skinny":*

In a second segment, *"Go skinny"*, we draw attention to the use of *gestures* and the *body* as common resources and how their use serves, like talk, to co-constitute a space and aid navigation.

In this second segment, another participant pair, George (VIP) and Sara (G), have just paid at a supermarket checkout and are heading towards the store's exit (Segment 2). Approaching a narrow gap to their right, they walk at a regular pace, almost side by side, with George holding Sara's arm (Figure 4a). As they near the gap, Sara prolongs the word *"because"*. She then pauses before saying *"go skinny"* (line 3). As they step forward, now at a slower pace, Sara stretches out her arm and then brings it behind her back. George is guided by Sara's arm movement and steps behind and further from her (Figure 4c).

As with the example above, this segment illustrates how pairs arrange their bodies and talk as they move through space and encounter obstacles. Thanks to Sara's gesture (stretching and moving her arm behind her), the two change their body configuration to pass through the narrow gap.

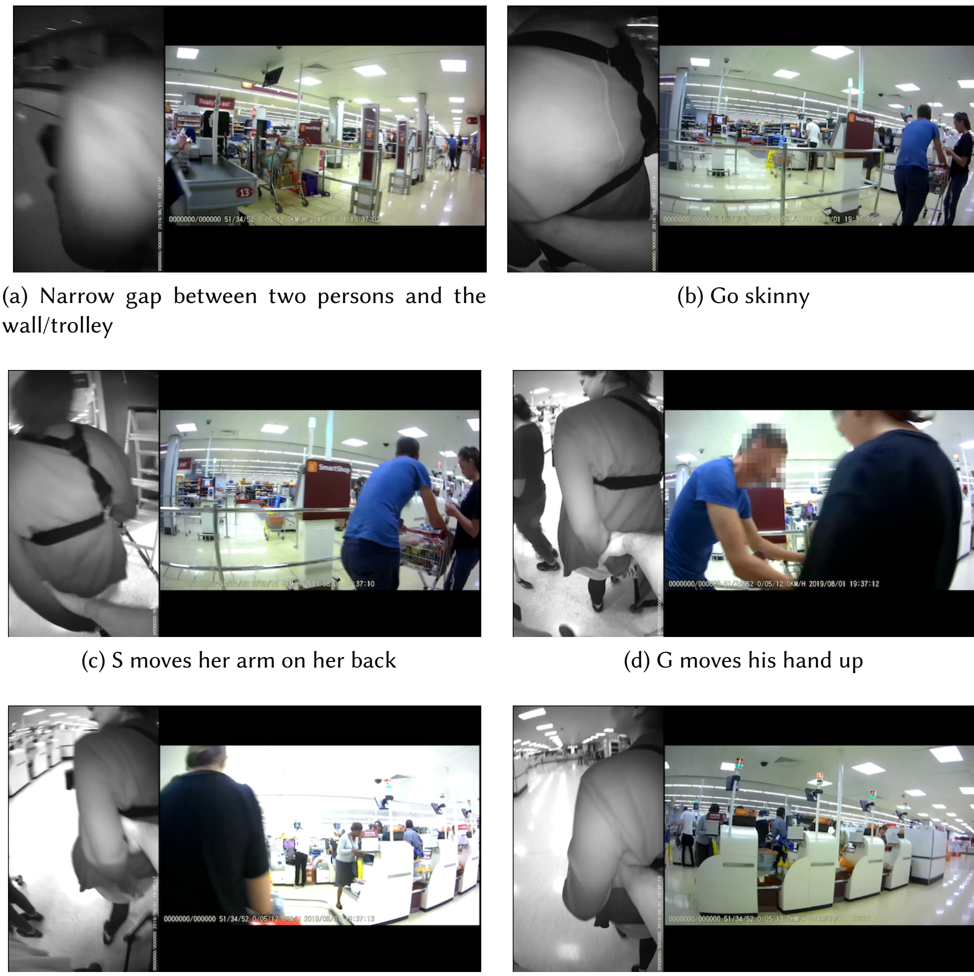
Sara's extended *"because"* and then *"go skinny"* anticipates this gesture, as does the pair's slowing down. Again, there is no need to explicitly announce the presence of an obstacle. Rather, the gap is indexed through the coordinated acts of speech and bodily movements.

Of particular interest in this segment is what follows. As the pair move through the gap, a pause in talk occurs. Not unlike the occasional pauses in talk between automobile drivers and their passengers [33], the pause marks a shift in focus and a tacit agreement that another matter demands immediate attention—the silence between George and Sara at one at the same time is demanded by the need to navigate the narrow gap and is a signal of the work they are engaged in to move through it. Moments later, the pair's resumption of talk is tacitly accomplished in a similar way. Moving his hand from Sara's wrist to her upper arm, George both returns to talking and moves to restore the side-by-side configuration (Fig. 4d and 4e). It is only then that Sara bends her arm to accommodate George beside her again (Fig. 4f). This sequence ends with a *"well done"* from George (line 4), a typical utterance used in closing a sequence [58] and Sara returning to the faster pace of walking.

What is apparent in this interchange is how it is not just that bodies and talk are working together. A common space is also being established, one where through a series of interwoven utterances and bodily movements a narrow gap is collectively established and navigated. Despite their different capacities for seeing and a variety of bodily and spoken interactions, full with nuance and subtlety, the two succeed in composing a space to move through. As Goodwin exemplifies in the analysis of talk between Chill, a man with aphasia, and his family, situations are made mutually intelligible through the situated conduct of the interlocutors (i.e., talk and gestures) [20]. Likewise, between George and Sara, space is made mutually intelligible and indeed actionable through the couples combined and sequential interactions. In other words, it is through their actions together, that a space-in-common is constituted.

Segment 2 - **Go skinny** (George and Sara)

```
                 Fig 4a
                 ▼
1        S:      she might not, she she was talking about it before she ran away
2        G:      right

                                                Fig 4b
                                                ▼
3  →     S:      becau::::se [pause] right go skinny

                 Fig 4c
                 ▼
                 [Sara moves her arm on her back. George's hand follows her arm. They are further
                 to each other. Slow down and Pause in talk]
                 0.8

                                                Fig 4d          Fig 4e
                 [George moves his arm from Sara's ▼rist to Sara's u▼per arm]

                 Fig 4f
                 ▼
4        G:      well // done [they speed up]
5        S:      *sh*thank you she was ( ) she was talking about before she went away becau::se she
6                said even though she spent a week with him she hasn't spent any time with ( ) she
7                is not gonna seen him for a couple of weeks [pause] she wants to [pause] see him
8                basically
9        G:      I am sure she will be fine
```



(a) Narrow gap between two persons and the wall/trolley

(b) Go skinny

(c) S moves her arm on her back

(d) G moves his hand up

(e) G holds S's upper arm

(f) G and S speed up

Fig. 4. Segment 2 - Go skinny (George and Sara)

In sum, through the above, we have illustrated how a common space can be co-constituted between people with vision impairments and their guides. That is, when couples do not share dominant sensory modalities (in our cases, sight), a particular work is made apparent that helps to establish a space in common and the capacity for movement through it. As we have seen, multiple resources are employed to provide information about bodies and their movements *vis-à-vis* the physical environment. Critically, this work is mutual, drawing on shared orientations to and mutual interchanges around talk, body gestures and movements, and objects.

4.2 Interdependence Interwoven into the Sequential Interactions between Pairs

The analysis so far highlights how resources do not always explicitly refer to the physical environment or deliberate navigation. Instead, the segments illustrate a complexity to the timings and rhythms people deploy while they coordinate their (inter)actions and interact with the environment and other objects.

Thus, we begin to see how an interdependence is enacted through a continuous, intermingling of the "simplest systematics" of talk and interaction [54]. The conventional idea of a guide chiefly doing the work of guiding a person with visual impairments belies the continual back and forth between the two actors, a relational achievement that makes even the most routine and unremarkable movements a highly collaborative venture. Although we want to avoid any crude parallels, we find a similar sophistication between people with vision impairments and their dogs [55], and indeed between people and dogs more generally [19, 38]. The point here is not that human guides are analogous to guide-dogs, but that there is a nuanced and interwoven character of such relations that turn on different but also shared sensory modalities and resources.

Following on from the above examples, through two further segments we want to further draw out the subtleties of this collaborative work and in doing so give particular emphasis to the interwoven and mutual engagements between people with visual impairments and their guides. That is, a necessary interdependence in their relationships.

4.2.1 Passers-by Approach: Turning to the third of our segments, where Alan (VIP) and Nick (G) are making their way to a Tube station, we find much in common with Segment 2, above. As in the example with Sara and George, who pause their conversation to pass through a narrow gap, we observe Alan and Nick momentarily falling silent as they navigate around some approaching passers-by (line 6-7). The conversation pauses as the pair slow down and veer to the right to avoid the upcoming pedestrians (see Figure 5). However, in contrast with our earlier example, there is no explicit announcement from Nick, of *"go skinny"* or otherwise. The change in direction and speed, and pause in talk appear to be sufficient for Alan to recognise that the space ahead is changing. Along with these cues, Alan responds accordingly, working in synchrony to maintain the silence and giving the pair the chance to attend to the approaching obstacle. We know that such synchronous interactions between interlocutors emerge through "a rich interplay between language processes and outward action" [51, p. 76], and that even the organisation and sequence of pauses and the return to talk depend heavily on verbal and nonverbal cues by both speakers [13]. This exchange then demonstrates a mutual dependence. Certainly, Nick is leading, but Alan is responsive to the subtle pace and rhythm of the cues and plays his role in constituting and navigating the space. If the situation was otherwise, we might imagine spoken interruptions, the need for repair, or possibly a collision with the passing pedestrians.

The same coordination is seen in the couple's return to their original walking pace and conversation. Guiding, Nick appears to initiate the increased pace in walking and as this occurs Alan restores the conversation (line 7), picking up where they left off. It is in this sense that the two are co-participants in the accomplishment of the navigation. Both are enabling each other through the use of verbal, bodily and spatial resources and through a mutually coordinated sequence of interactions.

Segment 3 - **Passers-by approach** (Alan and Nick)

```
      start  [Alan and Nick are walking at normal speed, next to each other. Alan is holding
             Nick's elbow]
  1   N:     yeah yeah did he recommend as an organization?
  2   A:     no::, I just I kinda did my research and I found it // and I was like oh god this
  3          is- this is what I found and it was like awesome=
  4   N:     *ah*
             0.2
  5   N:     =did he enjoy working for them?
  6   A:     I think he did yeah
             Fig 5
             ▼
  →          0.5 [Nick is slowing down while he is turning on his right and is passing close to
             other pedestrians and then speeds up again]
  7   A:     I think he- I think he knew he might interview me also // he is quite handy=
  8   N:     *ah, okay*
  9   N:     =( ) inside the track
 10   A:     yeah, pretty much
```

Fig. 5. Segment 3 - Passers-by approach (Alan and Nick)

4.2.2 Freezing-up: Through this next example, we want to continue with this idea of how a pair enables each other in and through their turn-by-turn interactions. The situation we now turn to, is though, a more complex one where a couple lose contact with one another. In the segment in question (Segment 4), Luke (VIP) and Alice (G) have just bought drinks and food in a café and are walking away from the till. As Luke takes the opportunity to move his cane from one arm to the other (Figure 6b), Alice lets go of him and then turns back to retrieve something from the counter (Figures 5c and 5d). At first glance, Luke's reaction to being left on his own, stopping suddenly, or *"freezing-up"*, looks to signal his inability to act. We might assume he is at a loss, unable to manage on his own.

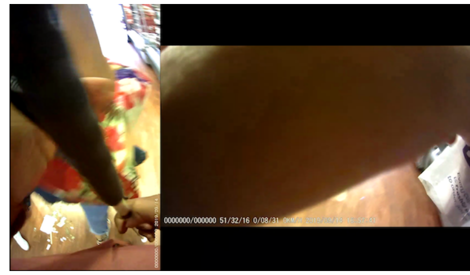
The reader should know, however, that Luke is in fact a highly independent man who lives on his own and regularly travels by himself. A more nuanced reading of the situation we encounter in Segment 4 is then that, in unexpectedly losing hold of Alice, Luke is responding to a *re*-configuration of space. As we have seen in each of the examples above, when a person with a vision impairment holds their guide, a space—a space to move in—is constituted together. To freeze here is to adjust to a new space and the alternative possibilities it affords. It is likely a response to avoid potential collisions with others, yet it also invites something more. As Middleton and Byles suggest in their studies of people with vision impairments travelling in cities, to freeze can also be a sign of one being open to other "interdependent exchanges" [43, p. 82]. Thus, Luke's freezing-up might be read equally as an opening to Alice to replace her arm, an active cue for her to re-establish physical contact.

Let us review the sequence (lines 3-6) in finer detail to consider the possibility of this greater interdependence between the pair. Luke's first action (moving his cane from one arm to the other, see Figure 6b) suggests he is already seeking to repair an unusual arrangement. In the sighted guide configuration they regularly adopt, Luke usually holds Alice's left arm. When Luke lets Alice's right arm go, he is thus seeking to re-establish their regular respective positions. Stepping away (Fig. 5c) and turning towards the counter (Fig. 5d), Alice appears to respond to Luke in stretching her left arm toward him (Fig. 5d). Not able to reach him, she says *"come on Luke"* (line 6), re-establishing a common space before she is actually able to make physical contact.

Segment 4 - **Freezing-up** (Luke and Alice)

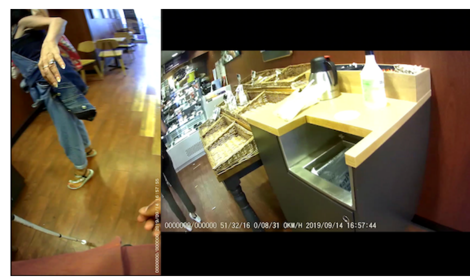
```
              [L and A they have just bought some food and they are approaching some table to
              have a seat]
1     A:      come one // Luke
2     L:      *do you want* to try the vegan

                  Fig 6a
                  ▼
3     A:      yeah

              Fig 6b
              ▼
  →           [L lets go A's arm to change the arm that supports the cane]

                                              Fig 3c
                                              ▼
              [L slightly stretches his arm to probably find Alice's left arm]
              0.4

                                          Fig 5d
                                          ▼
4     A:      ok yes, of course I forgot=
5     L:      =eh?
              0.2

                          Fig 5e
                          ▼
6     A:      come on Luke
```

(a) L and A are walking towards some chair

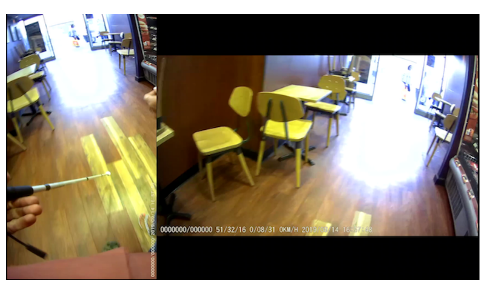


(b) L changes hand's cane

(c) A steps forward



(d) A turns toward the counter



(e) L is holding A's left arm

Fig. 5. Segment 4 - Freezing-up (Luke and Alice)

The broader point to draw from these segments is the ongoing and orchestrated work of interdependent interactions. At a micro-level of interaction, involving forms of talk, gestures and body movements, pairs build up and continually attune their relations with one another. For Alan and Nick, and Luke and Alice, the interdependencies between the couples move fluidly, but there is an agency in both directions, the actions from one member of a pair invite actions from the other, the resources are used by one and in turn create the conditions for the other to act. In this way, interdependence is threaded through the relations, it is an integral feature for couples moving in and through spaces together.

4.3 Ruptures and Repairs to Common Space

In their article *Troublesome Objects* [16], Due and Lange describe how the sweeping white cane can help to detect obstacles, but also simultaneously marks an arch in front of the person using it, alerting passers-by that the space should not be obstructed (also see [69]). Again, we suggest that this constituting of space is performed in a similar way between people with visual impairments and their guides. Between the pairs, we find a common space being co-constituted through fine-grained and nuanced interactions, and, specifically, emerging through interwoven and sequential acts of talk, bodily movements and gestures, and mutual references to objects. In this way, the interdependent production and use of space can be understood as a taken for granted feature of ordinary (inter)actions.

So far, we have presumed a fluid and untroubled co-production of these acts and the corresponding realisation of a common space. When navigating together, however, we also find pairs may need to put more explicit effort into establishing what, exactly, constitutes a common space and how to (inter)act in it. As we saw in the last example (4.2.2), ruptures can arise that demand repair to reestablish a space in common. In this section, our interest thus turns to how people with visual impairments and their sighted guides negotiate their ideas of space. Specifically, we present two segments, *"Wait here, wait here"* and *"Two for six"*. The former explores, in detail, how the coordinated actions of *"letting go"* can change the focus from a common space to a more personal one. The latter, *"Two for six"*, describes how a change in pitch is used as a resource to repair and re-establish a common space.

4.3.1 "Wait Here, Wait Here": In the first of these two examples, we return to Alan (VIP) and Nick (G), see Segment 5. Here, the pair are leaving a café, walking side-by-side with Alan holding Nick's left elbow. Reaching the doorway, Nick realises they have forgotten Alan's backpack. *"Wait here, wait here"*, he exclaims (lines 5-7), as he turns and removes his arm from Alan (Fig. 6a). Nick's utterance is pronounced using a firmer tone of voice to emphasise its different, instructional status in talk. As we see Nick return to the table where the bag has been left, Alan's subsequent question: *"that's bad, isn't it"* is left unanswered (lines 8-9).

We see here how a pair moves from coordinating their actions and co-constituting a common space, to operating alone. That is, the pair go from being physically co-proximate (with Alan holding Nick's arm), moving together, and Nick offering a verbal indication of his relative movement, to each individual operating in separate, personal spaces. Like Luke and Alice in the previous example (Segment 4), this transition from being together to being separated is not without its troubles. Alan's unanswered question suggests a problem: even though Nick alerts him with *"wait here"*, the follow-up question shows the transition is not mutually intelligible—that the changing circumstances are not understood simultaneously or equally by the pair. Alan's question is left as a rhetorical statement, whether intended or not. The contrast with the coordinated actions and flow of talk we see while pairs co-constitute a common space is stark.

The problem appears to turn on the abrupt shift between common and individual spaces. In her study of "interactional spaces", Mondada [44] details the systematic use of movements, gaze, body orientation and mutual adjustments to describe the transitions from passing pedestrians, unknown to each other, to their focused co-participation in public space. What Mondada demonstrates is the nuanced work that is performed to accomplish such transitions. Although the change is in the opposite direction—from shared to individual spaces—we might expect to see a similar work between Alan and Nick. And yet such work is notably absent. To begin, their physical contact serves as an explicit resource to share and negotiate space together—in the guiding configuration each individual's movement has a direct consequence on the other's. It would seem though that the hurried letting go is a trigger for the difficulties. Nick's ambiguous "wait here" and imperceptibly timed move away from Alan breaches the co-constituted space. Though sequential turns in talk can go some way towards easing the transitions and repairing them (again, see Segment 4), Alan's question fails to achieve this and we witness a rupture that goes unresolved, at least for a time.

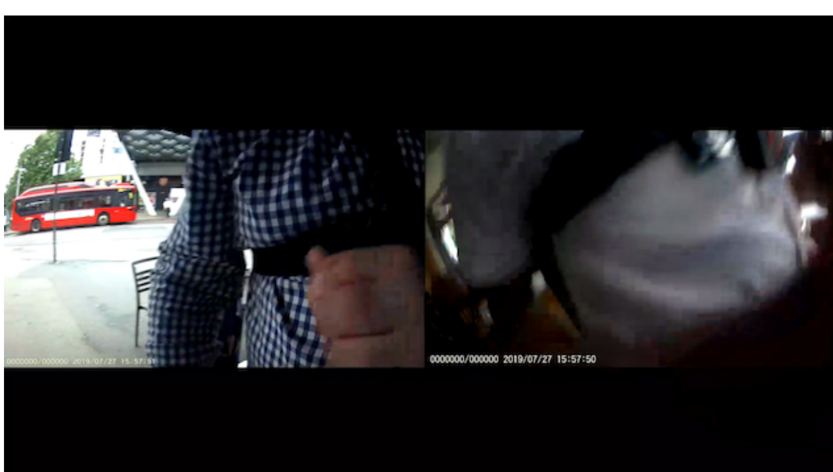
Segment 5 - **Wait here, wait here** (Alan and Nick)

1 N: alright
2 0.2
3 N: is it quite handy doing it here at the cafe that's give us another two minutes walk
4 A: yeah

Fig 6a ▼
5 ⟶ N: I left your bag behind, wait here
[they are in front of each other. A is not holding N's elbow]
6 A: oh //shit
7 N: *wait here*
[N goes back inside the cafe to retrieve the bag]

Fig 6b ▼
8 A: that's bad, isn't it?
9 0.3

Fig 6c ▼
10 A: haha::ha=
11 N: =ok I'll carry it, here it is
12 A: oh thanks man
13 N: hahaha



(a) N spontaneously turns toward A and the physical contact is broken

Fig. 6. Segment 5 - Wait here, wait here (Alan and Nick)

(b) N has left A's immediate proximity

(c) A laughs while N approaches him, lifting up his left arm

Fig. 6. Segment 5 - Wait here, wait here (Alan and Nick)

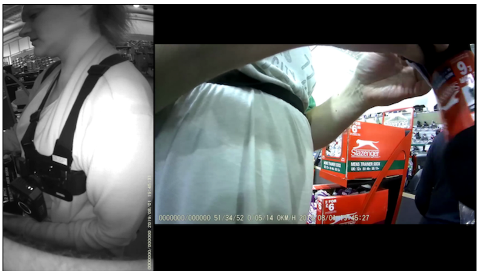
4.3.2 "Two for Six": In this next example, let us further examine the transition between common and personal spaces, and consider in particular how ruptures in the former can be open to repair. Sara (G) and George (VIP) are at a shop's checkout waiting to pay for some socks for George (see Segment 6). They are standing a short distance from one another in a queue, with George holding the pack of socks they have chosen earlier. Looking at and referring to the pack, Sara says *"this is four pounds"*. Although they are not in contact, the socks serve as something in common, an object they are mutually attending to through both touch and speech.

In the midst of this exchange, Sara notices a new pack of socks and decides to return the old ones to some hangers in a nearby aisle (line 6). Saying *"just put this back"*, she walks away from George (Figure 7d). The change in body configuration and distance marks a transition in their respective positions in space, though, as in the previous example, there are signs of trouble. The alert, *"just put this back"*, is ambiguous and the transition is not simultaneously recognisable to both parties. Indeed, Sara's utterance is said *sotte voce*, so may well not have been heard by George. Again, an interdependence between a pair is shown to be fragile and the common space enacted through them working in concert is ruptured. This is seemingly confirmed in George's next turn *"do you think"* (line 8). Although Sara has moved out of immediate earshot, George asks the question using the same pitch and without changing his orientation (Figures 7e and 7f), suggesting he is not aware Sara has left his immediate vicinity.

It is at this point that we see, on Sara's part, an attempt to repair the rupture in space. Saying *"You can get two for six pounds if you want"* (line 9), she raises the pitch of her voice and overlaps George's prior turn (line 8). Such overlaps have been noted as the source of troubles in video-mediated talk that is prone to latency and temporal delays [53], and are also known to be common in repairs to the order of turns between interlocutors [56]. Here, though, Sara appears to be making amends to a discontinuity in the co-constituted space between the pair. Her first *"...two for six"* utterance is a reference to the new pack of socks she has found, but in talking over George (seemingly with some urgency) she also makes apparent her change in location—possibly recognising he had not been aware of it. And, again, Sara's repetition of *"two for six pounds"* (line 11) is in response to George's questioning *"ehm?"*, but also reinstates her spatial location beside him and a return to a common space. Sara's words, then, do more than maintain the sequence of turns, they indicate both her spatial location and her return to the mutuality or interdependence between the pair.

Segment 6 - **Two for six** (George and Sara)

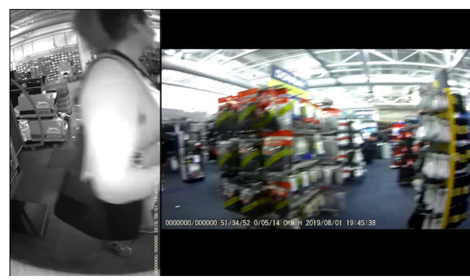
```
                                 Fig 7a
                                 ▼
 1        S:    this is four pounds
 2        S:    do you wanna get those actually?
 3        G:    yeah probably they will be alright=
 4        S:    =it's quite nice

                                            Fig 7b Fig 7c
                                            ▼      ▼
 5        G:    ( there's any one more once ) but::
                0.2

                               Fig 7d
                               ▼
 6   →    S:    ↓just put it back
                0.2
 7        G:    they are alright

                                 Fig 7e
                                 ▼
 8        G:    do you think e::hm they're normal-

                Fig 7f                                        Fig 7g
                ▼                                             ▼
 9        S:    ↑↑you can get two for six pounds if you want
                0.6
10        G:    ehm?=

                           Fig 7h
                           ▼
11        S:    =↑you can get two for six pounds
12        G:    do I need to
```



(a) S gives to G a new pack of socks

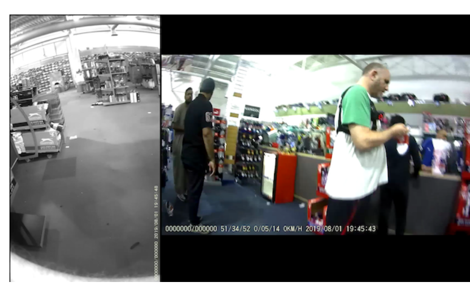
(b) S is putting the old pack of socks in a wrong place

(c) S turns on her left, she wants to put back the old pair of socks

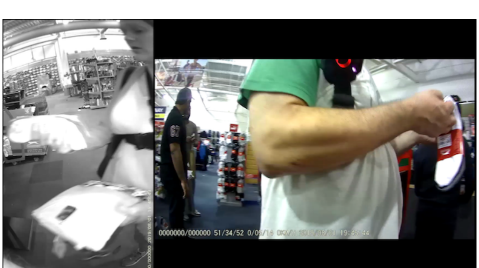


(d) S is going away from G to put back the old pack of socks

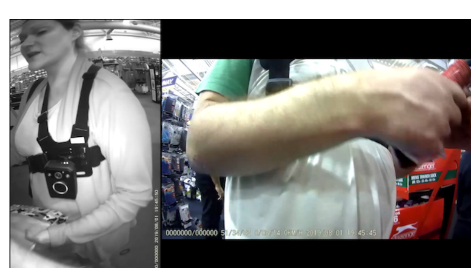
(e) G stands at the check out, while S is away



(f) S is coming back to the check out



(g) S is close to G



(h) S is repeating what she just said looking at G

Fig. 7. Segment 6 - Two for six (George and Sara)

In sum, in this last of our empirical sections we have seen how ruptures can occur in the mutual constitution of common space. In other words, the interdependent relations formed in and through navigation—between blind people and their guides—can be subject to breakdown, and such breakdowns can breach or rupture the co-constitution of common space. It would seem such ruptures occur when pairs lose the ability or fail to maintain the coordination of actions through talk, gestures and body movements, or a reference to common objects. We have also seen, however, that ruptures can be repaired through these same resources. For example, modifications to the sequential order of talk (e.g., 4.3.2) can be used to reconstitute a common space that has momentarily been disrupted. As we will see below, a recognition of such ruptures and repairs invites an opportunity for thinking about the role of AI in assistive technologies that support navigation.

5 DISCUSSION

In the above detailed examination of six segments from a larger corpus of video data, we have given particular focus to the ways people with vision impairments and their guides navigate together. Our findings, overall, show that this navigation involves a mutual investment in constituting a space to move through—*co-constituting a space in common*. Space here is understood not to precede interaction, nor is it a given. Rather it is jointly accomplished, brought about through an unfolding set of relations [14]. Step-by-step, turn-by-turn, and in close bodily correspondence with one another, the pairs above show that space is (and in some cases isn't) *made* navigable, together. It is, in this way, that we find an interdependence at work in the sighted guide relationship. For pairs, a back-and-forth in talk, bodily gestures and movements, and mutual references to objects are all involved in co-constituting and navigating a common space.

In this discussion, we use this perspective on the practical accomplishment of interdependent navigation to identify and discuss opportunities for AI-enabled AT.

Again, we wish to consider how AI might be employed not as a means to "solve" undertakings like navigation—that are so often situationally dependent and thus a significant challenge to model [62]—but to compliment and possibly extend individual and collective capacities [5, 6]. Our interest in AI is consequently not to make up for some deficiency in sighted guiding, but to consider what AI might offer in the collaborative achievements between actors. We seek to follow an all together more fundamental line of inquiry tied to questions about the role of AI in supporting collective autonomy and agency. It is through this line of inquiry that we point to, in the following subsections (see 5.1-3), three areas that might be considered in designing AI-enabled assistive technology. We bring the discussion to a close by reflecting on the limitations of the research we have conducted and our plans for future work.

5.1 The Use of Talk, Bodies and Objects as Resources

As we noted in Section 2, a common approach to aid navigation with AT is to view it as a problem of travelling from one place to another, using verbal or acoustic feedback to signal landmarks and ensure obstacles are avoided. A presumption is that navigation can be aided by recovering the *de facto* details of the physical environment. This arguably seeks to tackle "the problem" of navigation in sighted terms, privileging the visual organisation of a scene and providing a means for a user to "see" what is around them.

What we wish to highlight first, then, is that this framing of the problem may be misjudged. As an alternative, we begin from the basis that sighted guide partnerships are a routine part of navigating for people who are blind or vision impaired (and incidentally they are also deeply intertwined with friendships and family relations [18, 32]). A goal might thus be to support these partnerships rather than assume they can be replaced with technology. Our findings extend this perspective, suggesting that the details about an environment might be more usefully represented in terms of how they relate to one another in terms of temporal concurrence or sequence. This reduces emphasis on doing the work of recognition and, instead, places it on how, exactly, resources like *talk*, *objects*, and *body movements and gesture* are routinely used between people to navigate together.

In 4.1, we illustrated how such resources are used to co-constitute a space in common. Crucially, it is not the resources alone that help but how they are situated with respect to the sequential order of events. Thus the parsimonious use of language or gestures—e.g., saying *"step down"* or *"go skinny"*, or bending an arm—can provide the necessary detail that avoids long-winded disruptions and helps to co-configure a legible space for pairs to move in. Similarly, the pack of socks that George and Sara hold (4.3.2) does not present itself as a resource because of what it is, but how it is held by the pair and communicates their spatial relationship.

Our above examples also reveal how multiple, temporally concurrent actions and resources come to be critical to the ways pairs navigate together. Alice's *"Come on step down"* foreshadows the curb, but it is also set alongside Luke's cane sweeping across the curb (4.1.1). The resources, produced in concert as well as in sequence, are what attune the pair to their surroundings and, again, make features relevant so that they can co-constitute a space and navigate the curb without noticeable hesitation.

From this perspective, we suggest assistive technology that supports navigation could go beyond the mere description of the physical environment (e.g., the detection of specific obstacles nearby, and turn-by-turn instructions). An AI-enabled AT could approximate the salience of features in a scene by attending to how they are used as resources *vis-á-vis* the unfolding sequence of interactions. The key emphasis of such an AT would need to shift away from natural language processing, gesture or object recognition to mapping how relevant aspects such as talk, objects movements and gestures mutually give rise to their use as resources. For example, it is remarkable in the sequences between George and Sara (4.1.2) and Alan and Nick (4.2.1) that silence operates as a marker for approaching obstacles. The silence says, literally, nothing, but through its timing with approaching obstacles signals a necessary shift in attention between the pairs.

One area in which this idea of reorienting research to detect sequences and multiple, concurrent interactions may have material impact is on how AI-enabled systems are trained. Currently, ATs providing navigation support use computer vision and machine learning algorithms for object detection and recognition. These systems rely on trained models using well-known datasets such as *ImageNet* and *Microsoft's COCO*. Such large-scale datasets contain many labelled common objects placed in everyday settings, and modelling is often intentionally designed to complicate the recognition task and background/foreground segmentation. Generally, the goal is to extract the object from its surroundings and achieve a high accuracy in object recognition under varying conditions (including where the background is dynamic, e.g. video). Consequently, trained models can yield impressive recognition results that are largely dependent on the dataset/training data. Given the significance of the production of resources in sequence and concert, the challenge here may not be to extract details such as body parts and objects (and also talk and audio) from segmented backgrounds. Instead, it may be to determine ways of labelling data and producing datasets that account for how these details are placed in sequence or used in concert with what might ordinarily be treated as extraneous background noise. This would be to apply a greater attention to how features are *made* relationally relevant in contexts like sighted guide navigation. Our intention here is not to discount the considerable labour involved in data labelling and producing datasets, but to suggest there may be alternatives to mainstream paradigms in AI/ML recognition systems, ones that respond more directly to the constituted settings in which the systems are used, and are sensitive to the ways in which multiple resources are coordinated in action.

5.2 Greater Interdependencies

Research surrounding navigation-based AT might then benefit from shifting attention away from the problem of recognising details in a user's environment, to supporting the means by which the world is made meaningful through talk and interaction. Relatedly, such research might question a further presumption: that *independence* is the sought after goal, that people with vision impairments should, through technology, be given the means of navigating on their own. Again, in examples where AT has been designed to support navigation, it is often the case that it is treated as an individual accomplishment and thus something to be solved for the individual. Where the recognition of other people is addressed, it is either to recognise them as an obstacle [3] or to detail their attributes (e.g. "man, aged 35", etc.). What we want to suggest is that the support for navigation in assistive technologies might instead show a concern for how people actively work together, how they rely on one another to move together.

Consider the example *Freezing-up* (4.2.2) in which Luke seemingly freezes when left on his own. As we suggested, this might easily be read as him faltering, loosing his capacity to move as his guide, Alice, leaves his side to return to the counter in a café. However, it can at the same time be thought of as an opening. To *freeze* here is to open up the space for something else to happen, to create the conditions for another to act [43]—in this case for Alice to turn and step-back, and adjust to Luke's outstretched arm. The slowing down of pairs presents a particularly interesting case of coordinated actions. Across our examples, we repeatedly saw such slow-downs. We found them to be triggered by a guide's reduced walking pace, the movement of an elbow, or the resistance in forward momentum (and no doubt other hard to detect actions). And, they would often come with corresponding actions from the person being guided, a similar slow-down, and possibly a silent pause or step closer to the guide. As on other occasions, actions were reciprocated, but in this case they also made room for something more to happen. In our examples in sections 4.1.2 and 4.2.1, the change in walking pace accomplished between pairs created the conditions for attention to be given to other things, for bodies to be realigned, or talk to be resumed. These are small openings, but at the same time through a coordination between actors something new is made possible. The co-constituted space enacted between a pair affords new sets of further interaction. Many movements on the part of our participants might be understood in a similar way; slowing down, pauses in talk, interruptions and so on can be seen as ways to maintain a coherence in a pair's co-constituted common space.

Such micro-adjustments and subtlety in interaction present significant challenges for AI-enabled AT. Recognising how one person's actions trigger or are reciprocated by another—amidst a complex mixture of interactions and across highly variable contexts—is nontrivial and unlikely to be tractable by AI systems for some time, if at all. However, what might be manageable is the recognition of microlevel but detectable actions, and system output/feedback corresponding to these actions. So, what if an audible sound accompanied and reflected a pair's pace or even a pause in talk as they navigated an obstacle or approached passersby? Crucially, such interventions would not replace the coordination between pairs, as it were solving "the problem" of obstacles. Instead, it would aim to serve as a further resource for opening up the possibilities and potentially affording the space for new (inter)actions. Sound reflecting the presence and relative proximity of an obstacle during a pause in talk could, for instance, make the need to veer one way or another more predictable or a return in talk more open to both parties. These would though be resources that hold the options open rather than dictate specific actions. They would also need to be sensitive to the availability of information to both parties in ways that were not disruptive to ongoing interaction.

5.3 Ruptures and Repairs

The ruptures we recounted to common space (4.2.1 and 4.3.1-2) and the possibility for repair (4.3.2) present cases of both clear breakdowns and opportunities for intervention from AT. When pairs lose contact with one another, we find there can be an abrupt change to the space they are in. A shift must be made between a co-constituted and shared space to individually managed spaces, and problems can arise accordingly. In *"Wait Here, Wait Here"* (4.3.1), for example, Alan is left talking to himself when Nick lets go of his arm to retrieve the forgotten bag. The common "freezing" may be an action inviting another action, but it is always a reactive one and one that awaits a response from the guide or another bystander.

In conversation analysis, a *transition relevance place* refers to a moment in turn-taking where a place is opened for a follow-on turn from an interlocutor (i.e. when a transition to the next speaker becomes relevant) [54]. A question (ending with a raise in pitch) is one obvious example, but so to are pauses and non-linguistic utterances such as "huh", "uhmm" or "errr". To ease the abrupt ruptures to common space between people with vision impairments and their guides, we might imagine a similar cue provided by an AT. The challenge in designing the AT would be how to aid the noticing of mutual configurations of space. An AT could indicate a moment of *transition relevance* by signalling the possibility of a reconfiguration of space. Changes to actors' orientations to talk (e.g., moving from facing one another to looking outwards) or changes in their relative spatial arrangements (e.g., a shift in orientation to something else in space) could be relayed using audible or tactile feedback to indicate the possibility—the relevant place—for changes.

Most obvious here would be one member of a pair walking away. More subtle cues might also be recognised, however, such as a guide's abrupt rotation away from a pair's direction of travel, as in Alice's movement (4.3.2), or a guide's words said *sotto voce*, with indexical spatial references such as "over there" or "wait here" (e.g. 4.3.1 and 4.3.2). Further work would need to be done to understand how talk and bodies could be used as a resource in this way, but as above the key recommendation here is not to treat this as a replacement for existing cues, but as a way to add to the resources pairs already have available to them to make space mutually meaningful and relevant.

To extend this point, in another of our examples, *"Two for Six"* (4.3.2), Sara illustrates how subtle cues in interaction can repair breaches to a shared space. A spoken response to George's turn, and indeed a raise in pitch and the overlap in talk, signal both a change in the talk, but also serve as a bridge, reconstituting the shared space. We find another example of how the use of interactional resources—in this case talk and turn-taking—provides a means of managing space and bodily coordination. Yet, we also see that the recovery of ruptured space is, again, largely weighted towards action from the guide. When people with vision impairments and their guides are in contact there is a recognisable back and forth, each person may be capable in different ways but together they are able to coordinate their actions to move through space (for the most part) unproblematically. But when the common space is breached the balance changes, the guide becomes the proactive member of the pair; the person with the vision impairment becomes far more reactive in what they can do. *"Two for Six"* (4.3.2) illustrates this. The opportunity to repair is open to Sara as the guide. Continuing with our proposals for AT that compliment pairs' interactions, we might here consider how information is provided to re-centre the agencies, to allow more possibilities from both parties.

We can learn lessons here from the broader body of work in CSCW. For example, in their study of systems designed to assist driving, Perterer et al. [48] show how advanced driver assistance systems fail to make use of the collaborative work involved in driving, where front-seat passengers often come to share tasks and duties. The authors suggest that driving could be thought of in more distributed terms, and correspondingly in-car systems could be designed to further extend the joint work, involving front-seat passengers in activities such as monitoring the speed of the car and assisting with navigation devices. This resonates with Bennett et al.'s [5] discussion of "crowd work" in AT, in which systems might inspire new forms of collaboration where people with disabilities are not only the recipients of assistance but become proactive companions in completing tasks.

Returning to the presented research, such an orientation might be explored further by considering how bodies are located in space. The relative spatial representation of nearby people with respect to a user has been proposed elsewhere (e.g. [1, 21]). Here, though, a particular investment could be put into exploring ways to track the relative location and distance between a user and a guide. For instance, AI-enabled systems may provide information about a guide's proximity when letting go occurs. Computer vision and machine learning techniques could be adopted to identify the guide and provide information about orientation and relative distance between pairs. Feedback representing this information would provide a user with the resources to orient themselves towards a guide who has left their side and possibly even allow them to walk together without physical contact. This proposal is again tentative, but shows our aim to support the ongoing and emerging relations between people with vision impairments and their guides, and presents a perspective that makes for richer interdependencies and an expansion of collective capacities.

5.4 Limitations and Future Work

The research we have presented has been conducted through audio-visual recordings of participant pairs, and made explicit use of body cameras, detailed transcription, and interaction analysis. Altogether these have offered a valuable method to highlight the complex interplay of talk, gestures and body movements, and other objects in sighted guiding. However, we recognise there are some technical, ethical and methodological limitations that arise in the research and that deserve reflection to improve future studies.

At a technical level, body cameras were a useful tool to capture participant interactions, moment by moment. However, our perspective was limited to the video frames provided from both cameras. These views sometimes did not record everything that might have been relevant, and were obviously unable to capture the degree of bodily and tactile contact between our pairs. Despite our study procedure—that included demonstrating how to wear the cameras, and trial runs with our participants to find the optimal arrangement of cameras (see Section 3.2)—such limitations are thus likely to have led to missed details. For instance, in some cases a participant's camera had been set up not to capture the torso or arms of the other member of the pair, so we were unable to see how movements and gestures were used. Likewise, the pressure of a grabbing hand or the tension in an arm, shoulder, or the movement of a body were undoubtedly useful resources between our pairs that we might have had some visual indication of, but we could not confirm. We recognise that all methods come with their limits so our recommendation here is not to suggest additional recording equipment, but rather to recommend explicit acknowledgement of what is and is not available to analysts through specific methods (see below).

Body cameras also raised some ethical considerations throughout the study. Our research was conducted in public spaces and participants had cameras at their homes for 3 days. This raised the risk of data beaches, and consequently the danger of data access by unauthorized people. It was for this reason that we decided to use body cameras that encrypted media at the point of capture, with access only possible using a password. We are aware, however, that more precautions could have been considered with respect to the confidentiality of data. Capturing private and sensitive information is highly likely using video recordings and this may in turn increase the risk of harm for participants. We might have, for example, reduced such risks by asking participants to review their media and to delete any sensitive video after recording their journey and before researchers could access them.

The last consideration relates to our methodology and the relationship between the recordings and analysis, and the activity as experienced by participants. In research that uses video and conducts interaction analysis on such recordings, particular assumptions are made about the access to the situational details. In practice, this method allowed us to examine, in detail, a very particular form of social activity (namely, how social order is accomplished). However, it is evident that what we as analysts see and hear may not be what participants hear, see and, indeed, feel [30]. We are also likely to miss much of the intimacy and care that play a relevant role in pairs' relationships, a commitment to understanding one another and the compulsion to do well together [6]. It is fair to say that the method we chose authorised a particular understanding of sighted guiding, but may in doing so have risked glossing over the skill and know-how *felt* by our participants. We therefore recognise that other subtle dimensions of interaction may have been precluded in our method and thus were not available to us in our analysis, results and findings.

Our research provides a step toward future work on investigating and designing AI-enabled assistive technology that extends the ways companions walk and navigate together. Specifically, our own future work will focus on situations in which ruptures occur. We are currently investigating in more detail how such ruptures happen when pairs *"let go"* of one another. As above, our aim is to draw attention to the use of situational resources used by pairs in sighted guide partnerships. We intend to design and evaluate AI-based interventions based on this work and hope to demonstrate that AI can have a role in easing the transitions between common and personal spaces, and potentially extending the ways common spaces are co-constituted.

6 CONCLUSION

Our study gathered video recordings of 4 people with visual impairments working with their sighted guides to navigate during routine journeys. We analysed 40 segments of video in detail and used 6 of these to illustrate the collaborative and interdependent work involved in navigation.

Our findings reveal how people with visual impairments and sighted guides use multiple resources, such as *talk*, *body gestures and movements*, and *objects* to co-constitute a common space that can be navigated together. These resources do not always explicitly describe the physical environment, but they are employed by people to inform how to move their bodies in relation to one another and to space. Here, timings and rhythms play an important role to coordinate their (inter)actions. Additionally, we showed the interdependence work during navigation and the continuous shifting between agencies. Sighted guides and guided people with visual impairments actively take and respond to actions through resources, building up and negotiating a common space of interactions and understanding. This interdependent production allows them to accomplish navigation successfully. Finally, our findings also depict moments of rupture. These ruptures occur when people fail in the negotiation and coordination of actions through multiple resources, leading them to deploy talk, gestures and body movements, and objects to repair such breakdowns.

We have demonstrated that there is a rich tapestry of highly collaborative work and mutual agency in sighted guiding that constitutes a prime example of interdependence. Taking this perspective allows us to question current approaches to AI-enabled assistive technology. In particular we suggest a reorientation towards (1) augmenting the sighted guiding relationship rather than replacing it; (2) focusing on identifying the mutual use and sequence of talk, bodies and objects as resources that constitute a common space rather simply providing a description of the environment; (3) identifying new opportunities for AI-enabled interventions that complement the existing resources; and (4) enable and support repair activities when interaction ruptures are detected.

Our work provides a first step toward a new framing of AI-based assistive technology which recognises interdependence as central to the autonomy and agency of people with vision impairments.

ACKNOWLEDGEMENTS

We would like to thank all our participants for their enthusiasm and support throughout the study, and the anonymous reviewers for their detailed comments and suggestions. This work has been supported through a scholarship awarded by the doctoral training programme in the School of Mathematics, Computer Science and Engineering at City, University of London.

REFERENCES

[1] Subeida Ahmed, Harshadha Balasubramanian, Simone Stumpf, Cecily Morrison, Abigail Sellen, and Martin Grayson. 2020. Investigating the Intelligibility of a Computer Vision System for Blind Users. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 419–429. https://doi.org/10.1145/3377325.3377508

[2] Dragan Ahmetovic, Valeria Alampi, Cristian Bernareggi, Andrea Gerino, and Sergio Mascetti. 2017. Math Melodiess: Supporting Visually Impaired Primary School Students in Learning Math. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work - W4A '17*. 1–2. https://doi.org/10.1145/3058555.3058583

[3] Dragan Ahmetovic, Masayuki Murata, Cole Gleason, Erin Brady, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. Achieving Practical and Accurate Indoor Navigation for People with Visual Impairments. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work - W4A '17*. 1–10. https://doi.org/10.1145/3058555.3058560

[4] Mara Balestrini, Paul Marshall, Raymundo Cornejo, Monica Tentori, Jon Bird, and Yvonne Rogers. 2016. Joke box: Coordinating shared encounters in public spaces. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* 27 (2016), 38–49. https://doi.org/10.1145/2818048.2835203

[5] Cynthia Bennett, Erin Brady, and Stacy M Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *ASSETS*, Vol. 18. https://doi.org/10.1145/3234695.3236348

[6] Cynthia L Bennett, Daniela K Rosner, and Alex S Taylor. 2020. The Care Work of Access. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[7] Stuart Blume, Vasilis Galis, and Andrés Valderrama Pineda. 2014. Introduction: STS and Disability. *Science Technology and Human Values* 39, 1 (2014), 98–104. https://doi.org/10.1177/0162243913513643

[8] Stacy M Branham and Shaun K Kane. 2015. Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. 2373–2382. https://doi.org/10.1145/2702123.2702511

[9] Stacy M. Branham and Shaun K. Kane. 2015. The invisible work of accessibility: How blind employees manage accessibility in mixed-ability workplaces. *ASSETS 2015 - Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (2015), 163–171. https://doi.org/10.1145/2700648.2809864

[10] Barry Brown and Matthew Chalmers. 2003. Tourism and mobile technology. *Ecscw 2003* September (2003), 335–354. https://doi.org/10.1007/978-94-010-0068-0_18

[11] Barry Brown, Moira McGregor, and Donald McMillan. 2015. Searchable Objects: Search in Everyday Conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 508–517. https://doi.org/10.1145/2675133.2675206

[12] Michele A. Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P. Bigham, and Amy Hurst. 2012. Crowdsourcing subjective fashion advice using vizwiz: Challenges and opportunities. *ASSETS'12 - Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (2012), 135–142. https://doi.org/10.1145/2384916.2384941

[13] Joseph N Cappella and Sally Planalp. 1981. Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research* 7, 2 (1981), 117–132.

[14] Paul Dourish. 2006. Re-Space-Ing Place: "Place" and "Space" Ten Years On. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada) *(CSCW '06)*. Association for Computing Machinery, New York, NY, USA, 299–308. https://doi.org/10.1145/1180875.1180921

[15] Brian Due and Simon Lange. 2018. Semiotic resources for navigation: A video ethnographic study of blind people's uses of the white cane and a guide dog for navigating in urban areas. *Semiotica* 2018, 222 (2018), 287–312. https://doi.org/10.1515/sem-2016-0196

[16] Brian L. Due and Simon Bierring Lange. 2018. Troublesome Objects: Unpacking Ocular-Centrism in Urban Environments by Studying Blind Navigation Using Video Ethnography and Ethnomethodology. *Sociological Research Online* 24, 4 (2018), 475–495. https://doi.org/10.1177/1360780418811963

[17] Be My Eyes. 2017. *BeMyEyes*. https://www.bemyeyes.com/ Retrieved 15-May-2020.

[18] Kelly Fritsch. 2010. Intimate assemblages: Disability, intercorporeality, and the labour of attendant care. *Critical Disability Discourses/Discours critiques dans le champ du handicap* 2 (2010), 1–14.

[19] David Goode. 2007. *Playing with my dog Katie: an ethnomethodological study of dog-human interaction.* Purdue University Press.

[20] Charles Goodwin. 2004. A Competent Speaker Who Can't Speak: The Social Life of Aphasia. *Journal of Linguistic Anthropology* 14, 2 (dec 2004), 151–170. https://doi.org/10.1525/jlin.2004.14.2.151

[21] Martin Grayson, Anja Thieme, Rita Marques, Daniela Massiceti, Ed Cutrell, and Cecily Morrison. 2020. A Dynamic AI System for Extending the Capabilities of Blind People. 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, 1–4. https://doi.org/10.1145/3334480.3383142

[22] João Guerreiro, Dragan Ahmetovic, Daisuke Sato, Kris Kitani, and Chieko Asakawa. 2019. Airport accessibility and navigation assistance for people with visual impairments. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–14. https://doi.org/10.1145/3290605.3300246

[23] João Guerreiro, Jefrey P. Bigham, Daisuke Sato, Chieko Asakawa, Hernisa Kacorri, Edward Cutrell, and Dragan Ahmetovic. 2019. Hacking blind navigation. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–8. https://doi.org/10.1145/3290607.3299015

[24] J. M. Hans du Buf, João Barroso, João M.F. Rodrigues, Hugo Paredes, Miguel Farrajota, Hugo Fernandes, João José, Victor Teixeira, and Mário Saleiro. 2011. The SmartVision navigation prototype for blind users. *International Journal of Digital Content Technology and its Applications* 5, 5 (2011), 351–361. https://doi.org/10.4156/jdcta.vol5.issue5.39

[25] Christian Heath. 2011. Embodied action: Video and the analysis of social interaction. *Qualitative research* 3 (2011), 250–270.

[26] Christian Heath and Jon Hindmarsh. 2002. *Analysing Interaction: Video, ethnography and situated conduct.* Technical Report. King's college London, London.

[27] Christian Heath, Hubert Knoblauch, and Paul Luff. 2000. Technology and social interaction: the emergence of 'workplace studies'. *The British Journal of Sociology* 51, 2 (2000), 299–320. https://doi.org/10.1111/j.1468-4446.2000.00299.x

[28] Christian Heath, Marcus Sanchez Svensson, Jon Hindmarsh, Paul Luff, and Dirk Vom Lehn. 2002. Configuring awareness. *Computer Supported Cooperative Work (CSCW)* 11, 3-4 (2002), 317–347.

[29] Rabia Jafri and Marwa Mahmoud Khan. 2018. User-centered design of a depth data based obstacle detection and avoidance system for the visually impaired. *Human-centric Computing and Information Sciences* 8, 1 (dec 2018), 14. https://doi.org/10.1186/s13673-018-0134-9

[30] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences* 4, 1 (1995), 39–103. https://doi.org/10.1207/s15327809jls0401_2

[31] Seita Kayukawa, Keita Higuchi, João Guerreiro, Shigeo Morishima, Yoichi Sato, Kris Kitani, and Chieko Asakawa. 2019. BBEEP: A sonic collision avoidance system for blind travellers and nearby pedestrians. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–12. https://doi.org/10.1145/3290605.3300282

[32] Christine Kelly. 2013. Building bridges with accessible care: Disability studies, feminist care scholarship, and beyond. *Hypatia* 28, 4 (2013), 784–800. https://doi.org/10.1111/j.1527-2001.2012.01310.x

[33] Eric Laurier, Hayden Lorimer, Barry Brown, Owain Jones, Oskar Juhlin, Allyson Noble, Mark Perry, Daniele Pica, Philippe Sormani, Ignaz Strebel, Laurel Swan, Alex Taylor, Laura Watts, and Alexandra Weilenmann. 2008. Driving and 'passengering': Notes on the ordinary organization of car travel. *Mobilities* 3, 1 (2008), 1–23. https://doi.org/10.1080/17450100701797273

[34] Young Hoon Lee and Gérard Medioni. 2016. RGB-D camera based wearable navigation system for the visually impaired. *Computer Vision and Image Understanding* 149 (2016), 3–20. https://doi.org/10.1016/j.cviu.2016.03.019

[35] Eric Livingston. 1987. *Making sense of ethnomethodology.* Taylor & Francis.

[36] BeSpecular LTD. 2017. *BeSpecular.* https://www.bespecular.com/ Retrieved 15-May-2020.

[37] Hannah Macpherson. 2017. Walkers with visual-impairments in the British countryside: Picturesque legacies, collective enjoyments and well-being benefits. *Journal of Rural Studies* 51 (2017), 251–258. https://doi.org/10.1016/j.jrurstud.2016.10.001

[38] Clara Mancini, Janet van der Linden, Jon Bryan, and Andrew Stuart. 2012. Exploring Interspecies Sensemaking: Dog Tracking Semiotics and Multispecies Ethnography. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 143–152. https://doi.org/10.1145/2370216.2370239

[39] Parth Mehta, Pavas Kant, Poojan Shah, and Anil K Roy. 2011. VI-Navi : A Novel Indoor Navigation System for Visually Impaired People. *International Conference on Computer Systems and Technologies – CompSysTech'11 VI-Navi:* 11 (2011), 365–371. https://doi.org/10.1145/2023607.2023669

[40] Helena M. Mentis and Alex S. Taylor. 2013. Imaging the body: Embodied vision in minimally invasive surgery. *Conference on Human Factors in Computing Systems - Proceedings* (2013), 1479–1488. https://doi.org/10.1145/2470654.2466197

[41] Microsoft. 2019. *SeeingAI.* https://www.microsoft.com/en-us/ai/seeing-ai Retrieved 15-May-2020.

[42] Microsoft and Guide Dogs for the Blind. 2014. *The Cities Unlocked.* https://www.guidedogs.org.uk/cities-unlocked Retrieved 15-May-2020.

[43] Jennie Middleton and Hari Byles. 2019. Interdependent temporalities and the everyday mobilities of visually impaired young people. *Geoforum* 102, April (2019), 76–85. https://doi.org/10.1016/j.geoforum.2019.03.018

[44] Lorenza Mondada. 2008. Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics* 41 (2008). https://doi.org/10.1016/j.pragma.2008.09.019

[45] Ingunn Moser. 2006. Disability and the promises of technology: Technology, subjectivity and embodiment within an order of the normal. *Information Communication and Society* 9, 3 (2006), 373–395. https://doi.org/10.1080/13691180600751348

[46] Optelec. 2014. *Ruby - Handheld Magnifier.* https://uk.optelec.com/products/880123-007-ruby-hd.html Retrieved 15-May-2020.

[47] World Health Organization. 2020. *WHO - Global Data on VI.* https://www.who.int/blindness/publications/globaldata/en/ Retrieved 15-May-2020.

[48] Nicole Perterer, Petra Sundström, Alexander Meschtscherjakov, David Wilfinger, and Manfred Tscheligi. 2013. Come drive with me: An ethnographic study of driver-passenger pairs to inform future in-car assistance. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (2013), 1539–1548. https://doi.org/10.1145/2441776.2441952

[49] Beryl Plimmer, Andrew Crossan, Stephen A Brewster, and Rachel Blagojevic. 2008. Multimodal collaborative handwriting training for visually-impaired people. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. 393. https://doi.org/10.1145/1357054.1357119

[50] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–12. https://doi.org/10.1145/3173574.3174214

[51] Daniel Richardson, Rick Dale, and Kevin Shockley. 2008. Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. *Embodied communication in humans and machines* (2008), 75–94.

[52] RNIB. 2020. *Sighted Guide Technique.* https://www.rnib.org.uk/advice/guiding-blind-or-partially-sighted-person Retrieved 15-May-2020.

[53] Karen Ruhleder and Brigitte Jordan. 2001. Co-constructing non-mutual realities: Delay-generated trouble in distributed interaction. *Computer Supported Cooperative Work (CSCW)* 10, 1 (2001), 113–138.

[54] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4 (dec 1974), 696. https://doi.org/10.2307/412243

[55] Clinton R Sanders. 2000. The impact of guide dogs on the identity of people with visual impairments. *Anthrozoös* 13, 3 (2000), 131–139.

[56] Emanuel A Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in society* 29, 1 (2000), 1–63.

[57] Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Linguistic Society of America* 53, 2 (1977), 361–382.

[58] Emanuel A Schegloff and Harvey Sacks. 1973. *Opening up Closings.* Technical Report.

[59] Kristen Shinohara, Jacob Wobbrock, and Jacob O Wobbrock. 2011. In the shadow of misperception: Assistive technology use and social interactions Aligned Rank Transform View project Supporting navigation for blind people View project In the Shadow of Misperception: Assistive Technology Use and Social Interactions. (2011). https://doi.org/10.1145/1978942.1979044

[60] Kristen Shinohara and Jacob O Wobbrock. 2016. Self-Conscious or Self-Confident? A Diary Study Conceptualizing the Social Accessibility of Assistive Technology. *ACM Transactions on Accessible Computing* 8, 2 (2016), 1–31. https://doi.org/10.1145/2827857

[61] British Blind Sport. 2020. *Sport activities.* https://britishblindsport.org.uk/play-sport/links/ Retrieved 18-Sep-2020.

[62] Lucy Suchman. 2007. *Human–Machine Reconfigurations: Plans and Situated Actions.* Cambridge University Press. https://doi.org/10.1017/cbo9780511808418.017

[63] M. Swobodzinski and Raubal M. 2009. An Indoor Routing Algorithm for the Blind: Development and Comparison to a Routing Algorithm for the Sighted. *International Journal of Geographical Information Science* 23, 10 (2009), 1315–1343.

[64] Anja Thieme, Cynthia L Bennett, Cecily Morrison, Edward Cutrell, and Alex S Taylor. 2018. "I can do everything but see!"-How People with Vision Impairments Negotiate their Abilities in Social Contexts. (2018). https://doi.org/10.1145/3173574.3173777

[65] Anja Thieme, Cecily Morrison, Nicolas Villar, Martin Grayson, and Siân Lindley. 2017. Enabling Collaboration in Learning Computer Programing Inclusive of Children with Vision Impairments. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. 739–752. https://doi.org/10.1145/3064663.3064689

[66] Yingli Tian, Xiaodong Yang, Chucai Yi, and Aries Arditi. 2013. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Machine Vision and Applications* 24, 3 (2013), 521–535. https://doi.org/10.1007/s00138-012-0431-7

[67] Ryan Wedoff, Lindsay Ball, Amelia Wang, Yi Xuan Khoo, Lauren Lieberman, and Kyle Rector. 2019. Virtual showdown: An accessible virtual reality game with scaffolds for youth with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[68] Peter Weeks. 2013. Synchrony Lost , Synchrony Regained : The Achievement of Musical Co-Ordination Author(s): Peter Weeks Reviewed work (s): Published by : Springer Stable URL : http://www.jstor.org/stable/20011105 . Your use of the JSTOR archive indicates your acceptan. 19, 2 (2013), 199–228.

[69] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. "just let the cane hit it": how the blind and sighted see navigation differently. In *ACM SIGACCESS conference on Computers & accessibility*, Vol. 8. 217–224. https://doi.org/10.1145/2661334.2661380

[70] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–14. https://doi.org/10.1145/3290605.3300341

[71] Michael Zöllner, Stephan Huber, Hans Christian Jetter, and Harald Reiterer. 2011. NAVI - A proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6949 LNCS, PART 4 (2011), 584–587. https://doi.org/10.1007/978-3-642-23768-3_88

A INTERACTION ANALYSIS SYMBOL LEGEND

- = Latching, no interval between the end of a prior and start of the next piece of talk.
- \- A single dash indicates a cut off either because of an interruption or self-repair.
- // Overlapping, the double oblique indicates the point at which a current speaker's talk is overlapped by the talk of another.
- ∗text∗ Text between asterisks indicates what has been said by another speaker during overlapping.
- : Colon(s) indicate that the prior syllable is prolonged. Multiple colons (e.g., :::) indicate a more prolonged syllable.
- ↑ An upward arrow indicates a marked rise in pitch.
- ↓ A downward arrow indicates a marked lowering of pitch.
- [] Squared brackets are used to describe what is happening visually, such as movements, speed, gesture and so on.
- ( ) Single pairs of parentheses indicate that words are unclear or inaudible in the clip.
- (( )) For vocalisations which are not easy to spell out such as ((cough)), ((snort)), and ((sniff)).
- <u>text</u> Underlined text indicates a different voice tone.
- → Points to the location of the phenomenon being discussed.
- 0.5 Indicates time in seconds between two 2 turns talking.

Received June 2020; revised October 2020; accepted December 2020