Smartphone-Based Assistance for Blind People to Stand in Lines

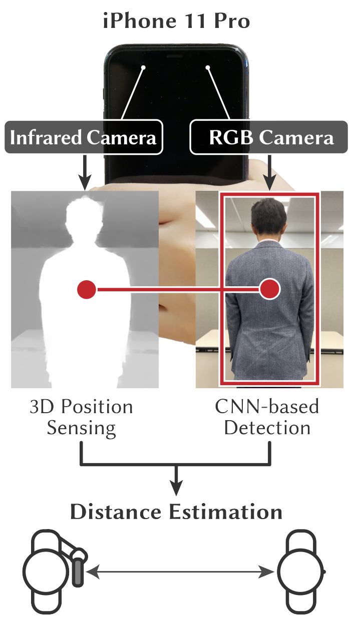

Figure 1: The pedestrian detection and the distance estimation via an off-the-shelf smartphone.

Seita Kayukawa
Waseda University
Tokyo, Japan
k940805k@ruri.waseda.jp

Shigeo Morishima
Waseda Research Institute for Science and Engineering
Tokyo, Japan
shigeo@waseda.jp

Hironobu Takagi
IBM Research - Tokyo
Tokyo, Japan
takagih@jp.ibm.com

Chieko Asakawa
IBM Research
Yorktown Heights, New York, United States
chiekoa@us.ibm.com

João Guerreiro
LaSIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal
jguerreiro@di.fc.ul.pt

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.
© 2020 Copyright is held by the author/owner(s).
ACM ISBN 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3382954

Abstract
We present a system to allow blind people to stand in line in public spaces by using an off-the-shelf smartphone only. The technologies to navigate blind pedestrians in public spaces are rapidly improving, but tasks which require to understand surrounding people's behavior are still difficult to assist. Standing in line at shops, stations, and other crowded places is one of such tasks. Therefore, we developed a system to detect and notify the distance to a person in front continuously by using a smartphone with a RGB camera and an infrared depth sensor. The system alerts three levels of distance via vibration patterns to allow users to start/stop moving forward to the right position at the right timing. To evaluate the effectiveness of the system, we performed a study with six blind people. We observed that the system enables blind participants to stand in line successfully, while also gaining more confidence.

Author Keywords
Visual impairment; orientation and mobility; assistive technology; pedestrian detection; vibrotactile feedback.

CCS Concepts
•Human-centered computing → **Accessibility technologies; •Social and professional topics** → *People with disabilities;*

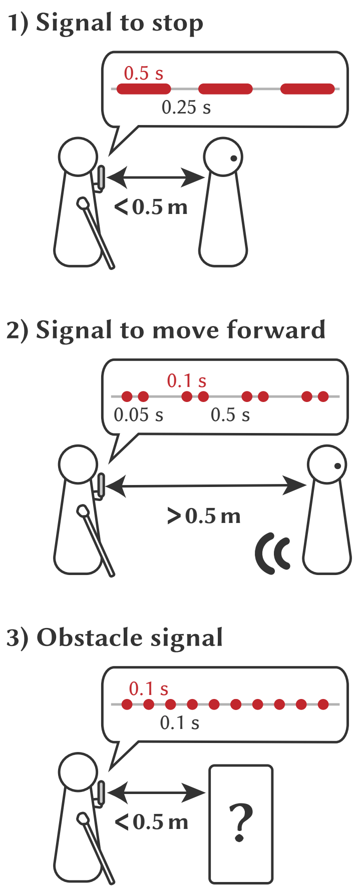

Figure 2: Policy of vibration emission.

Introduction

Standing in line is a challenging task for blind people as it requires an accurate perception not only of the distance to the person in front, but also of the timing and the direction that the line is moving. Still, it is a very important activity to participate in society: standing in line is required in public transportation and services, shops, cafes, among many other examples. Despite the prominence of research and solutions for navigation assistance, most efforts focus on turn-by-turn guidance or awareness of surrounding Points-of-Interest [5, 9, 17, 20] – both relying on static maps – or obstacle avoidance [4, 8, 10, 15, 21, 11]. However, standing in line requires a greater understanding of other people's behaviour, which is not supported by current solutions.

We present a system that enables blind people to stand in line in public spaces, using an off-the-shelf smartphone to detect and notify the distance to the person in front of the user. An iPhone 11 Pro is used to capture real-time depth data via a built-in infrared depth sensor (Figure 1) in addition to a standard RGB camera. The system uses RGB images for detecting pedestrians and the depth data to estimate distances. According to the sensing results, it provides three types of vibration alerts (Figure 2). Based on these alerts, blind users standing in line can stop when a person is within close distance, and sense movement to move forward to the *right position* and at the *right timing*.

To evaluate the effectiveness of our system, we performed a study with six blind users. We observed that the proposed system enabled participants to sense the line movement and to stand in line effectively by themselves. Moreover, participants felt more confident and comfortable to stand in line by themselves after the experiment. Based on our findings and user feedback, we discuss requirements to make the system practical and applicable for other use cases.

Related Work

Computer vision-based assistive technologies such as OrCam [1], Seeing AI[2], and Envision[3] are becoming increasingly popular among blind people. These systems are able to recognize printed letters, registered faces, and selected objects. However, they cannot recognize social context and as a consequence are not capable of assisting blind people to seamlessly participate in challenging social activities in public spaces, such as queuing. Other solutions are trying to make use of off-the-shelf smartphones to recognize obstacles in real-time – for instance using the Project Tango and smartphones with an infrared depth sensor [13, 8] – but the social context is also not addressed. Computer vision has been used to convey social behaviors to robots, including to stand in line [6, 14, 18]. Despite the prior work on robot navigation for blind people [12, 7], we found none capable of assisting blind people standing in line.

Proposed System

We developed a smartphone-based system that can detect surrounding people and inform about the distance to the closest person (Figure 1). This system intends to complement blind users' orientation and mobility skills in a social context, allowing them to stand in lines by themselves.

We used an off-the-shelf smartphone, iPhone 11 Pro[4], which is equipped with RGB image sensors and an infrared depth sensor. The system first detects pedestrians from the RGB streams by using a vision-based object recognition engine, YOLOv3-tiny [16], which recognizes "human" as an object type. Then, it automatically lists all bounding boxes and selects the largest one as the target. The distance is

[1] https://www.orcam.com/
[2] https://www.microsoft.com/en-us/ai/seeing-ai
[3] https://www.letsenvision.com/
[4] https://www.apple.com/iphone-11-pro/

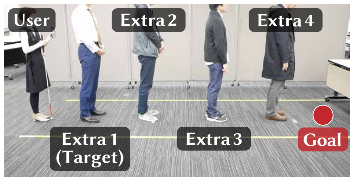

Figure 3: The start positions of each condition.

estimated as the depth data at the target central position (Figure 1). All these estimation steps were done at about 2–3*fps*. We checked that the system could estimate the distance between 0.2 m and 6 m with reasonable accuracy.

To convey distance information, we rely on vibration alerts as audio could be less effective in noisy, crowded environments [2]. The system emits three types of vibration alerts:

1. A **Signal to stop** indicates that a person is standing within 50 cm and that the user should stop moving. We used a long vibration alert (pulse duration (PD): 0.5 s and inter-pulse interval (IPI): 0.25 s) (Figure 2(1)).
2. A **Signal to move forward** indicates that a person is standing in front of the user, at a distance longer than 50 cm. The signal is used to recommend the user to step forward and uses a two-pulse vibration. (Figure 2(2)).
3. An **Obstacle signal** indicates an imminent risk of collision with any obstacle (pedestrian, desk, wall, etc.) located within 50 cm. We used a short vibration alert (the PD and the IPI: 0.1 s) as the signal (Figure 2(3)).

The absence of vibration indicates that the user lost the target person. In that case, users should scan the environment with their phone in order to find them. We set the vibration threshold to 50cm, because the distance of personal space while standing in line is around 40–80 cm [14] and users held the smartphone in front of them.

User Evaluation

To evaluate the effectiveness of our system, we performed a user study with 6 blind people (Table 1).

Tasks and Conditions

All tasks started with 5 people in line. The blind participants were asked to follow a line with four sighted people (*extras*) in front of them (Figure 3). They were asked to stand in line and proceed until reaching the desk reception (the goal). A researcher signaled for the *extra* standing in the first position to leave the line after 30, 60, or 90 seconds. Waiting times were randomized per *extra* and per trial. Each participant held the smartphone with one hand and used their white-cane on the other hand. We stopped the task if the participant overtook the person in front (*Extra 1* – referred as *target person* – in Figure 3). We designed two types of organized straight lines: in **C1**, four extras moved one by one. In **C2**, two extras in consecutive positions (randomly from 1 to 4 – Figure 3) left the line together. C2 was designed to evaluate response to irregular line movement. In order to simulate a crowded public space, we played ambient noise recorded at a shopping mall at 60 dB [19].

Procedure

We performed a pre-questionnaire where we asked participants about their prior experiences and challenges of standing in lines. We also asked them to rate a set of statements (Q1–Q5 in Figure 6) using 7-point Likert items (from 1: strongly disagree to 7: strongly agree). A training session of 10–15 minutes was then given to participants.

Then, participants performed six trials where they stood in line using our system until reaching the reception desk. The order of the line conditions was randomized for each participant (three C1, three C2). After the trials, participants were asked a set of questions for their confidence and comfortableness (Q1–Q5), the System Usability Scale (SUS) [3], and open-ended questions to gather qualitative feedback.

Metrics

In order to better understand how the line moved, we considered both timing and position. For timing, we measured the *reaction time*, which is the duration between the movement of an extra and the movement of the following extra

ID	Demographic info				Reaction Time (seconds)	SUS Score (Grade)	
	Age	Gender	Eyesight	Navigation Aid			
P1	22	Male	Blind	Cane	2.97 ± 0.68	**77.5**	**(B+)**
P2	33	Female	Blind	Cane	6.73 ± 5.48	37.5	F
P3	33	Female	Blind	Cane	4.39 ± 2.25	**80**	**(A-)**
P4	22	Male	Blind	Cane	3.50 ± 1.38	**87.5**	**(A+)**
P5	24	Male	Blind	Cane	2.43 ± 0.30	**97.5**	**(A+)**
P6	23	Male	Blind	Cane	2.18 ± 0.50	**90**	**(A+)**
				Average (Mean $\pm$ SD)	3.55 ± 2.66	**78.3 ± 21.3**	**(B+)**

Table 1: Participants' demographic information, their reaction time, and values for SUS scores.

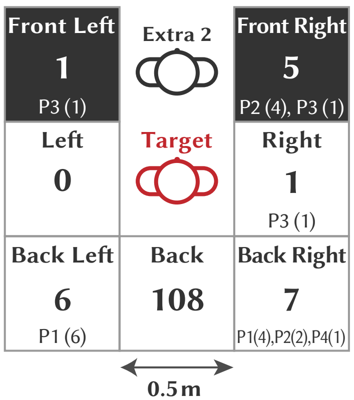

Figure 4: The distribution of positions where blind participants stopped.

or blind participant. We then compared the reaction time of the blind participants and of the sighted extras. For position, we defined seven positions around a target person as the *stop positions* of a participant (Figure 4). Ideally, the participant stops right after (*Back*) the Target, but a slight deviation to either side is also acceptable (Back Left or Right).

Results

Past Experiences and Opinions about Standing in Line
Participants reported standing in line in various situations, such as counters at stores, cafes, and airports, at a bus stop, and getting on the subway. Most participants reported trying to cope with standing in lines by using echolocation or their intuition with ambient sounds (P1, P2, and P6), asking people in line for help (P3), or touching clothes of the person in front (P5). P4 does not stand in line by himself.

Despite their strategies to stand in lines, all participants reported occasions where they did not realize the line was moving or bumped into a person in front of them: **A1:** *"In noisy places such as shopping centers and stations, it is hard for me to recognize when the line was moving. Even if I noticed the line movement, I can't sense the distance to the person in front."* (P6); and **A2:** *"Lines are not always straight. For example, when I stand in a serpentine line, it is difficult to determine the direction I should walk."* (P1).

Overall Performance
Figure 4 shows the distribution of positions where participants stopped after line movement. Overall, participants managed to stop just behind the target person (84.4%, 108 out of 128). We also noted that each blind participant tended to shift to a specific side during the tasks. The success rate to stop at the right position is 94.5% (121 out of 128) if considering slight deviations to the side. The task success rate for C1 and C2 tasks was 75% (18 out of 24), because we stopped a trial after one failure. Specifically, P3 and P2 overtook the target person two and four times, respectively. Figure 5 shows an example of a failure case. When the user shifted aside and lost the target person, the user tried to scan around to find the target person (Figure 5(1)). However, the system detected another person before detecting the target person (Figure 5(2)). As a result, the system recommended the user to move forward, even though the target person was standing on the users' left.

Table 1 reports the reaction time for each participant. The mean (M) reaction time was 3.55 seconds (SD: 2.66 sec.) for blind participants and 1.23 sec. (SD: 0.40 sec.) for the target person, significantly different by using Welch T-Test ($p < 0.001$, 95% CI: 0.95~1.49).

Subjective Ratings
Figure 6 shows the questionnaire results, where most participants felt more confident and comfortable to stand in line after the experiment (with the system) than before (without the system). In Q1–Q4, all participants except P3 in Q3 increased their scores after the experiment. Four participants (P2–P5) also increased the score of Q5. Table 1 reports the

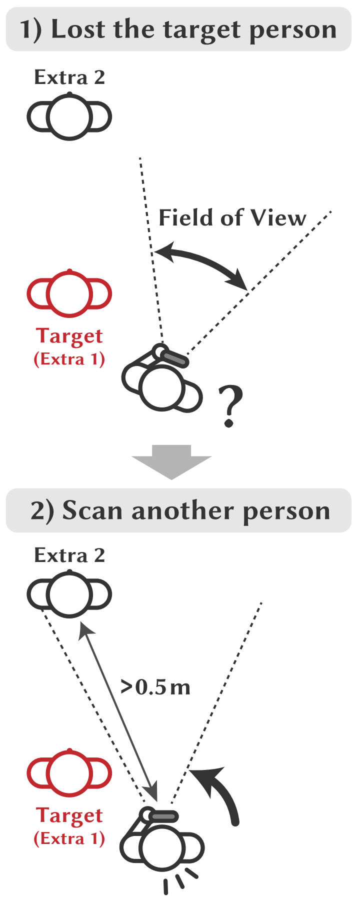

Figure 5: Example of failure case where a participant overtook the target person in front.

SUS scores [1] for each participant. The mean SUS score was 78.3 (SD: 21.3), which can be classified as "acceptable". P2 was the only participant with a lower score, mainly due to difficulties to hold the smartphone.

Qualitative feedback
Participants generally agreed that the system allowed users to start/stop moving forward at the right timing, as illustrated by their comments: **A3:** *"The biggest advantage of the system is that I could easily recognize the movement of a step forward from the person in front."* (P1); **A4:** *"By using the system, I could decide when and how far I should move forward."* (P4); and **A5:** *"The system provides the distance information, so it can reduce risks of collisions."* (P2).

Some participants provided positive feedback on our smartphone-based interface: **A6:** *"The system is implemented on the smartphone. It is a strong advantage of the system because I don't have to carry extra devices."*; and **A7:** *"The system was simple and easy to use."* (P4). In contrast, P2 commented that keeping the position of the system while waiting in line was difficult due to the large and heavy system: **A8:** *"This smartphone is big and heavy, so it was difficult for me to hold the smartphone stably."*.

The vibration alerts received positive feedback overall: **A9:** *"I could distinguish vibration patterns easily. I like tactile feedback more than audio because tactile-based alerts do not block ambient sounds."* (P2). Still, P3 suggested to use sound-based alerts rather than vibration: **A10:** *"It was a little difficult to distinguish between the three types of vibration. I think that using audio cues can be a good idea."*

When asked for suggestions, two users mentioned that the system should provide more detailed distance information or the direction information toward the target person: **A11:** *"When I lost the target person, I had to search him by myself while changing the direction of the system. I want to know which direction the target person stands before I lose the target person."* (P4); and **A12:** *"I would be happy if I can be aware of more detailed distance information. So, it can be a good option to change the pulse duration continuously to encode distance information."* (P1).

As another concern about our system, P2–P4 and P6 pointed out that they feel uncomfortable to point a smartphone to other people: **A13:** *"My concern is that others may be wondering why I'm pointing the smartphone at people."* (P3); and **A14:** *"To turn the touch screen to other people may seem strange to surrounding people."* (P6).

Discussion

All participants commented that they face difficulties to stand in line in their daily lives at subways, stores, cafes, and other public places even with their white cane and echolocation skills. In spite of such reality, the system successfully enabled participants to stand in line until reaching the desk independently. The system received 78.3 SUS mean score (B+, Acceptable), and all participants increased most of their scores for confidence and comfortableness (Q1 – Q5). The reaction time was significantly slower than that of sighted extras, so more studies are required to evaluate the acceptability of this latency in real-world situation. As for the appropriateness of the position, participants walked straight (Figure 4) in most cases. The task success rate for the organized line task was 75% and failure cases only happened for P1 and P2. Overall, these results indicate the potential of our approach, but also it suggests there is a large room for improvement.

The cause of all task failures was positioning, due to miss-tracking of the target person due to shifting aside. In order to observe shifting in harder situations, we briefly tested two

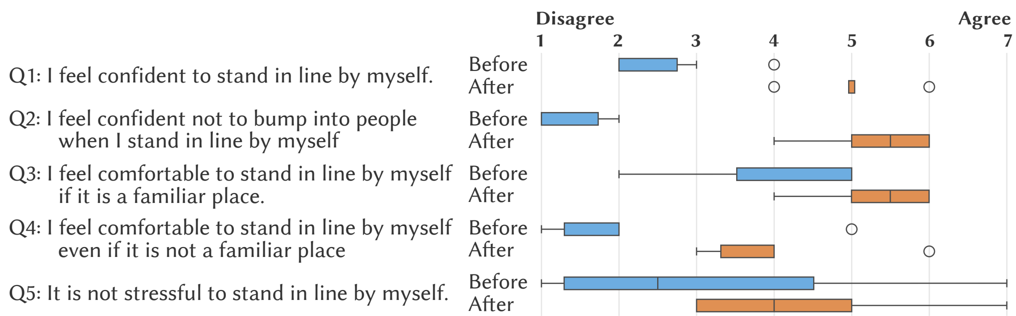

Figure 6: Likert items and a summary of the answers before/after the experiment.

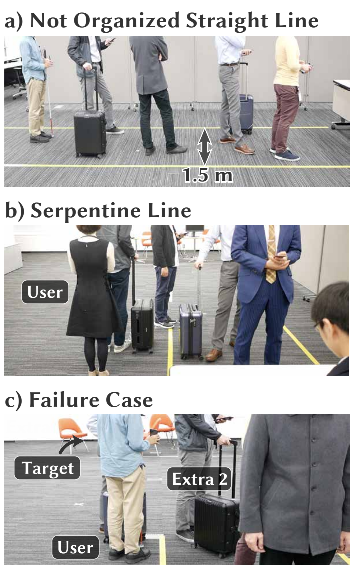

Figure 7: a) Non organized straight line, b) serpentine line, and c) a failure case where the user tracked a wrong person (Extra 2) rather than the target person.

more conditions (with all participants): 1) a non-organized straight line where *extras* stood in line within 1.5 m in width while carrying a suitcase, using a smartphone, or talking with others (Figure 7(a)); and 2) a serpentine line where *extras* would turn right to make a U-turn in the middle of the line (Figure 7(b)). All participants could finish the non-organized straight line without overtaking the target person. Also, all participants except P1 finished the serpentine line. As shown in Figure 7(c), P1 followed a wrong person who stood in front of the target person. A potential solution for improving target tracking is to identify a specific person as this would avoid recognizing and following the wrong person. This feature would also help assisting the user searching for the target after shifting aside and providing additional guidance. In addition, a greater understanding of the whole line could potentially assist users walking into a new line.

One limitation we found was the ergonomics of holding a smartphone during the task. While 5 out of 6 participants gave high SUS scores (77.5 – 97.5), P2 rated the system as “not acceptable” (37.5) mostly due to her difficulty to hold the smartphone despite trying a few variations of gripping and holding. As smartphones with advanced sensors are becoming larger and heavier, other alternatives should be investigated to ensure the applicability of solutions that rely on off-the-shelf smartphones. This is even more relevant if we consider that participants shared their concerns about the social acceptance of pointing a smartphone to another person. While this concern is shared with other vision-based technologies on smartphones, the close distance to the target person may amplify acceptability issues. Alternative design options, such as using a neck-strap to hold the smartphone at the chest-level may be more successful in seamlessly blending in the social context and reduce "social friction". In addition, activities that raise social awareness to the importance of vision-based technologies for blind people can help increasing their acceptability.

Conclusion

We presented a smartphone-based system that assists blind people to stand in lines. The system uses a built-in RGB camera and infrared camera of an off-the-shelf smartphone to estimate the distance to a person in front of the user. Based on the distance information provided by vibration alerts, blind users can stop or start moving forward to the right position at the right timing. We performed a study with six blind users, and results showed that our system allowed blind participants to stand in line successfully with increased confidence. We see a great potential to improve the independence of blind pedestrians in stand-in-line situations by making use of the sensing capabilities and processing power of off-the-shelf smartphones. However, further research is needed in order to improve target tracking and provide alternative designs that improve the acceptability of such approach in realistic, public scenarios.

Acknowledgements

This work was supported by JST ACCEL (JPMJAC1602) and JST-Mirai Program (JPMJMI19B2). We thank T. Ishihara, M. Murata, T. Itoh, and M. Oikawa for their support.

REFERENCES

[1] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[2] Nicholas A Bradley and Mark D Dunlop. 2002. Investigating context-aware clues to assist navigation for visually impaired people. In *Proc. of Workshop on Building Bridges: Interdisciplinary Context-Sensitive Computing, University of Glasgow*. 5–10.

[3] John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[4] Dimitrios Dakopoulos and Nikolaos G Bourbakis. 2010. Wearable obstacle avoidance electronic travel aids for blind: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1 (2010), 25–35. DOI: http://dx.doi.org/10.1109/TSMCC.2009.2021255

[5] Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. 2013. Indoor human navigation systems: A survey. *Interacting with Computers* 25, 1 (2013), 21–33. DOI: http://dx.doi.org/10.1093/iwc/iws010

[6] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. 2007. Natural person-following behavior for social robots. In *Proc. ACM/IEEE International Conference on Human Robot Interaction (HRI '07)*. ACM, 17–24. DOI:http://dx.doi.org/10.1145/1228716.1228720

[7] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M. Kitani, and Chieko Asakawa. 2019. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *Proc. ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. ACM, New York, NY, USA, 68–82. DOI: http://dx.doi.org/10.1145/3308561.3353771

[8] Rabia Jafri, Rodrigo Louzada Campos, Syed Abid Ali, and Hamid R. Arabnia. 2017. Visual and Infrared Sensor Data-Based Obstacle Detection for the Visually Impaired Using the Google Project Tango Tablet Development Kit and the Unity Engine. *IEEE Access* 6 (2017), 443–454. DOI: http://dx.doi.org/10.1109/ACCESS.2017.2766579

[9] Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Valeria Alampi, Hironobu Takagi, and Chieko Asakawa. 2018. Insights on Assistive Orientation and Mobility of People with Visual Impairment Based on Large-Scale Longitudinal Data. *ACM Transactions on Accessible Computing (TACCESS)* 11, 1 (2018), 5. DOI: http://dx.doi.org/10.1145/3178853

[10] Robert K Katzschmann, Brandon Araki, and Daniela Rus. 2018. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 3 (2018), 583–593. DOI: http://dx.doi.org/10.1109/TNSRE.2018.2800665

[11] 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. In *Proc. ACM CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 52:1–52:12. DOI:http://dx.doi.org/10.1145/3290605.3300282

[12] Diego López-de Ipiña, Tania Lorido, and Unai López. 2011. Indoor navigation and product recognition for blind people assisted shopping. In *Proc. International Workshop on Ambient Assisted Living (IWAAL '11)*. Springer, 33–40. DOI: http://dx.doi.org/10.1007/978-3-642-21303-8_5

[13] Eitan Marder-Eppstein. 2016. Project Tango. In *ACM SIGGRAPH 2016 Real-Time Live!* ACM, 25–25. DOI: http://dx.doi.org/10.1145/2933540.2933550

[14] Yasushi Nakauchi and Reid Simmons. 2002. A social robot that stands in line. *Autonomous Robots* 12, 3 (2002), 313–324. DOI: http://dx.doi.org/10.1023/A:1015273816637

[15] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle sonification for people with visual impairment or blindness. In *Proc. ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. ACM, 402–413. DOI: http://dx.doi.org/0.1145/3308561.3353779

[16] J. Redmon and A. Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[17] Timothy H Riehle, Shane M Anderson, Patrick A Lichter, Nicholas A Giudice, Suneel I Sheikh, Robert J Knuesel, Daniel T Kollmann, and Daniel S Hedin. 2012. Indoor magnetic navigation for the blind. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBS '12)*. IEEE, 1972–1975. DOI: http://dx.doi.org/10.1109/EMBC.2012.6346342

[18] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. 2015. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics* 7, 2 (2015), 137–153. DOI: http://dx.doi.org/10.1007/s12369-014-0251-1

[19] Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. 2016. *The geography of transport systems*. Routledge.

[20] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2019. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing (TACCESS)* 12, 3 (2019), 14. DOI: http://dx.doi.org/10.1145/3340319

[21] Limin Zeng, Markus Simros, and Gerhard Weber. 2017. Camera-based Mobile Electronic Travel Aids Support for Cognitive Mapping of Unknown Spaces. In *Proc. ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 8, 10 pages. DOI: http://dx.doi.org/10.1145/3098279.3098563