Investigating the Intelligibility of a Computer Vision System for Blind Users

Subeida Ahmed
Centre for HCI Design
City, University of London
London, United Kingdom
Subeida.Ahmed@city.ac.uk

Harshadha Balasubramanian
Department of Anthropology
University College London
London, United Kingdom
ucsahba@ucl.ac.uk

Simone Stumpf
Centre for HCI Design
City, University of London
London, United Kingdom
Simone.Stumpf.1@city.ac.uk

Cecily Morrison
Microsoft Research
Cambridge, United Kingdom
cecilym@microsoft.com

Abigail Sellen
Microsoft Research
Cambridge, United Kingdom
asellen@microsoft.com

Martin Grayson
Microsoft Research
Cambridge, United Kingdom
Martin.Grayson@microsoft.com

ABSTRACT

Computer vision systems to help blind users are becoming increasingly common, yet often these systems are not intelligible. Our work investigates the intelligibility of a wearable computer vision system to help blind users locate and identify people in their vicinity. Providing a continuous stream of information, this system allows us to explore intelligibility through interaction and instructions, going beyond studies of intelligibility that focus on explaining a decision a computer vision system might make. In a study with 13 blind users, we explored whether varying instructions (either basic or enhanced) about how the system worked would change blind users' experience of the system. We found offering a more detailed set of instructions did not affect how successful users were using the system nor their perceived workload. We did, however, find evidence of significant differences in what they knew about the system and they employed different, and potentially more effective, use strategies. Our findings have important implications for researchers and designers of computer vision systems for blind users, as well as more general implications for understanding what it means to make interactive computer vision systems intelligible.

CCS CONCEPTS

• Human-centered computing~Human computer interaction (HCI) • Human-centered computing~Accessibility • Computing methodologies~Computer vision

KEYWORDS

Computer vision, assistive technology, blind users, intelligibility.

ACM Reference format:

Subeida Ahmed, Harshadha Balasubramanian, Simone Stumpf, Cecily Morrison, Abigail Sellen, Martin Grayson. 2020. Investigating the Intelligibility of a Computer Vision System for Blind Users. In *Proceedings of ACM Intelligent User Interfaces (IUI'20). ACM, New York, NY, USA, 11 pages.* https://doi.org/10.1145/3377325.3377508

IUI '20, March 17–20, 2020, Cagliari, Italy

ACM ISBN 978-1-4503-7118-6/20/03…$15.00

https://doi.org/10.1145/3377325.3377508

1 Introduction

It is estimated that currently approximately 36 million people worldwide are blind [6]. To help them, advances in computer vision are being applied to assistive technologies, for example to recognize objects and describe scenes using their smartphones e.g. Seeing AI[1], TapTapSee[2], to help in navigating their environment [7,28], or to locate and identify people around them in a social situation [38,40,45].

However, in common with other AI systems, these computer vision systems targeted at blind users are 'black boxes' because their internal workings are unknown to the user [1]. There have been many recent calls to make AI systems transparent [9,42], alongside efforts to explain the functioning of computer vision systems to sighted users in many different domains, often using pixel-based visualizations [17,30,36]. We see *intelligibility* of an AI system as the ability of a user to build an appropriate mental model [31,32] that guides the user's interactions with the system. The user's mental model can be influenced through explicit explanations of the system's behavior [4,22], through instructions or tutorials [23], or simply through exploring and

[1] https://www.microsoft.com/en-us/ai/seeing-ai
[2] https://taptapseeapp.com/

interacting with a system [11,23]. Previous research [22,23,25] has indicated that explicit explanations about system decisions can lead to better understanding of how a system works, and that in turn, increased understanding leads to better user experience. Thus, understanding how this novel assistive technology operates is arguably crucial to their effective use by blind users.

Making computer vision systems that help blind users to navigate or explore their environment understandable faces two main challenges. First, current visual approaches to explain system behavior are inappropriate for blind users. Second, these systems usually provide a continuous stream of information, which makes explaining the decision-making of the system in real time problematic. How systems become 'transparent' or intelligible to sighted or blind users through other means, for example through exploration and interaction, or through instructions and tutorials, has received less attention.

Our aim was to study the intelligibility of a computer vision system that helps blind users locate and identify people in their vicinity using a continuous stream of information. In this case, users can only come to know this system through exploration or instruction. We were particularly interested in whether basic or enhanced instructions about how the system worked would affect the user experience of the system, reflect in users' interactions with the system, and their understanding of how the system worked. Our research questions were as follows:

- How do basic or enhanced instructions influence the user experience?
- How much, what type and what level of knowledge do blind users gain of the system?
- What is the effect of intelligibility on user interactions and use of the system?

Our work holds lessons for designers of intelligible computer vision systems for blind users, and contributes a deeper understanding of making AI systems intelligible for all users.

2 Related Work

Much recent work has focused on making AI system interpretable and transparent [2,3,8,9,26,29,34,43]. One way to make a system transparent is through explanations [14,22,25,27,39], and some work has provided (sighted) users with explanations of a computer vision system's behavior [17,18,20,30,36,44]. Explanations of how the system reasoned can lead to a deeper understanding of system actions and behavior [23] and better user experience by increasing user satisfaction [15,41], or user trust and/or reliance [5,10,33].

Less work has been conducted on exploring how systems become 'transparent' or intelligible to users through other means, for example through exploration and interaction, or through instructions and tutorials. The theory of mental models [31] argues that all users build an internal representation of how a system works through exploring a system as from previous interactions with similar systems, and then use this mental model to predict its actions and to shape their interactions and use with the system. We can distinguish different *types* of knowledge that might be encoded in a mental model: 'knowing that' and 'knowing how' [37]. Declarative knowledge ('knowing that') are facts and data whereas procedural knowledge ('knowing how') is associated with skills and rules [16]. Rasmussen's Skill-Rules-Knowledge (SRK) framework [35] adds that there might be different *levels* of mental models, for example, in a shallow (functional) mental model users employ a rule-based level, whereas in a deep (structural) mental model, users are able to employ a knowledge-based level.

3 Methods

The study employed a between-group design with two levels of instruction about how the system worked: basic intelligibility and enhanced intelligibility. Participants with visual impairments used a prototype system to carry out a semi-realistic task which mimicked networking at a career event. A mixed-methods approach was used which included both quantitative measures and qualitative data to gain a deeper understanding of the effects of intelligibility.

3.1 Prototype Overview

The prototype, developed by Microsoft Research, is comprised of a wearable headset (Figure 1) and a computer vision system which helps blind users to locate people around them, and to identify them if they have been trained into the system. The wearable headset has four cameras for detecting people, providing a detection angle of 160 degrees. The headset is capable of providing continuous spatialized audio, using a variety of sounds as auditory icons to feedback to the user about what it can 'see', and announcing identified individuals' names.

Figure 1: The Wearable Headset Prototype. It consists of an adjustable "headband" with one forward-facing camera, and three peripheral cameras. It also has bone-conducting earphones to provide spatialized audio. A small battery pack is included to the rear of the headset.

The computer vision system rests primarily on a pose and an ID model. The pose model recognizes that a person is present by detecting landmarks such as shoulders and a face. The ID model identifies an individual using facial recognition, in addition to building a world in which the movement of people is tracked.

3.2 Participants

Thirteen participants (12 males and 1 female) completed the study, ranging in age from 17 to 33 (*M* = 20.85). Participants were randomly allocated to either the Basic or Enhanced intelligibility group. Even though all participants were registered blind, there was a considerable variation in their visual abilities. Table 1 shows their background information, and their allocated condition.

ID	Age	Gender	Vision	Group
1	33	Male	No light perception for 10 years.	Basic
2	23	Female	Full field of vision, sees objects more than 6 meters away, since birth.	Basic
3	17	Male	Light perception, since birth.	Enhanced
4	17	Male	Light perception and full field of vision, since birth.	Enhanced
5	18	Male	Severely reduced field of vision, since birth.	Enhanced
6	22	Male	Light perception, sees objects more than 6 meters away, full field of vision, since birth.	Basic
7	21	Male	Light perception, sees objects between 3 and 6 meters away, since birth.	Enhanced
8	20	Male	Light perception, since birth.	Basic
9	18	Male	Light perception, sees objects more than 6 meters away, since birth.	Enhanced
10	20	Male	Light perception, sees objects more than 6 meters away, full field of vision, since birth.	Basic
11	20	Male	Light perception, sees objects between 3 and 6 meters away, since birth.	Enhanced
12	22	Male	Light perception, since birth.	Basic
13	20	Male	Light perception, sees objects between 3 and 6 meters away, full field of vision, since birth.	Enhanced

Table 1: Background and Allocated Group of Participants

The participants were recruited through VICTA, a charity running a week-long UK-based residential event in June 2019 for young people who are blind or partially sighted to learn technology and communication skills. None of the participants had previously encountered a wearable computer vision system such as the one used in our study. An ad was sent via email, and those who wished to take part received an accessible electronic consent form. If below the age of 18, the consent form was signed by a parent or guardian. Approval for this study was granted by the IRB board of Microsoft and noted by the Computer Science Ethics Committee of City, University of London. We did not pay any incentives but the participants received a few items of company merchandise, and detailed feedback about their communication and social skills.

3.3 Procedure

The study included a pre-task session to familiarize the participant with the system. The main session mimicked a career networking event where the participants used the prototype to find a recruitment specialist to talk to. The post-task session gathered responses from the participants about the use of the system and their background information. Overall, the whole session lasted for 45 minutes, and the participants completed the study individually.

3.3.1 Pre-task Session

The pre-task session stage was led by a sighted and a blind researcher who alternated facilitation.

When the participant arrived, they were told what the study would involve. They were then familiarized with how to wear the system and the system's basic functions, such as the volume control. The participants in the study were randomly allocated into two groups, the Basic Intelligibility group and the Enhanced intelligibility group. All 13 participants in the study received basic intelligibility instructions of the system, while seven participants were provided with further enhanced intelligibility instructions. The basic instructions explained what a specific audio sound meant; the sound was played to the participants to provide a better understanding for them:

- If you hear the "person identifier knock" sound, the system has detected that a person is in front of you.
- If you hear the "face identifier knock" sound, the system has not only detected that a person in front of you but has detected their face as well. If the system can identify who it is, it will announce the person's name.
- The "woodblock" sounds help you find a face. They get higher in pitch as the camera aligns with a person's face. The "snap" sound indicates that the system has found the person's face.

The enhanced instructions gave further information on the internal workings of the system:

- The system works within a certain range. It can detect a person up to 10 meters away. This is about the length of two cars or about 15 walking steps. It can identify people's faces best when they are within 4 meters, about slightly less than a car length or about 6 walking steps.
- The system will detect people within a 160-degree angle in front you. This is about shoulder to shoulder.
- To detect a person, it needs to see the head (eyes, nose, ears) or the torso (shoulder, chest and arms).
- The system needs to see the eyes, ears and nose of a person to positively identify him or her.

- It will read out the name of a recognized person each time your gaze crosses the midline of the person who is directly in front of you.
- The system remembers where people are and will assume that they stay in the same place for 10 seconds since it last saw them. Therefore, the system may read out a name despite the person not being in front of you anymore. It may also read out a name if a person is not looking at you, but still in front of you.
- There are a number of reasons the system could produce a 'person identifier knock' sound but not read out a name:
  - they are too far away.
  - their eyes, ears or nose are obscured or turned away.
  - the user is moving his or her head to quickly for a clear image.
- If you hear a 'person identifier knock' sound not followed by a name, the person is either not trained in the system or has not been recognized yet.

Last, the researchers reminded the participant of the name of the recruiter they were to look for in the career networking task, and they were guided to the networking event which was being conducted next door.

3.3.2 Main Session

We decided to use a career networking event for this study for three reasons: 1. A networking event is an example of a social context in which people with visual impairments would find the system useful, adding to the ecological validity of the study. 2. It encouraged the participant to use the features of the system to find and identify the recruiter. 3. We were able to control extraneous variables and ensure each session was repeatable.

Figure 2: Networking Task Setup During Main Session. Confederates talking to each other were situated in the middle of the room, with a technology demonstrator and a recruiter to the back of the room. The participant entered to the front-left, and needed to locate the recruiter to the back-right.

The career networking event was carefully orchestrated to mimic a real event yet make it repeatable (Figure 2). On entering the room, two study confederates were talking to each other in the middle of the room. If the participant started a conversation with them, they could answer his questions but otherwise not engage or point them in the direction of the recruiter. To the left at the back of the room was a demonstrator showing an accessible programming environment. To the right in the far corner away from the participant was the recruitment specialist, alongside some tables with refreshments. The recruiter was asked to avoid looking at the door when the participant came in, and during the first 7 minutes of the networking event remained seated silently in the back of the room so it would not make the task of finding her too easy. If the participant had spent over 7 minutes looking unsuccessfully for the recruiter, the recruiter would approach the participant. The participants were allowed to use their assistive technologies as per usual; sighted guides were instructed to ensure safety but not to lead.

In addition to the headset cameras, we also recorded the activity in the room with a static camera placed at the middle far end of the room.

3.3.3 Post-task Session

This part began with the NASA-TLX survey [13] to measure perceived workload for aspects of Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. We developed a 21-point tactile, high-visibility scale supplemented with braille stickers intended for the blind researcher to note down their answer.

The second part of the post-task session was an interview which measured the knowledge the user had of the system. We framed this as teaching someone else how to use the system. Participants were given the option of demonstrating their answer by using the system. We included several follow-up questions that probed a participant's understanding of the system further:

- If the headset calls out the name of a person, what can the user assume the system can see?
- If Tom was sitting down and the person he was trying to identify was standing up, where should Tom look to identify the person?'
- If the system played this (woodblock) sound, what should he do?
- If Tom took all the advice on board and could still not hear a name, what should Tom assume?

Finally, background information was collected using the World Health Organization's definition of severe sight impairment (blindness), and the Royal National Institute for the Blind's criteria for being certified severely visually-impaired.

3.4 Data Collection and Analysis

3.4.1 User Experience

To obtain measures for actual and perceived user experience, we used task success and workload ratings.

We used two measures of task success. First, we timed how long it took the 13 participants to find the recruiter from when they entered the room. We then performed a Mann-Whitney U Test to investigate any differences between the Basic and Enhanced group. Second, we analyzed the model predictions to measure how accurate the system was in identifying the recruiter. We calculated the accuracy as the fraction of times the system identified the recruiter as NEW or UNDETECTED until the correct ID. Unfortunately, the system only captured the data for six participants, two of which were in the Enhanced group, and thus our results are only indicative.

The responses to the NASA-TLX questions were entered in a spreadsheet as the participants answered. We translated the score given on the tactile scale to the raw workload scale between 0 to 100; a higher rating indicated more demand, thus a low rating is 'good'. A mental model score for one participant was unfortunately not entered; this data point was excluded from the analysis. A Mann-Whitney U test was performed to find out whether there was a significant difference in workload between the Basic Intelligibility group and the Enhanced intelligibility group. Sometimes, participants would mention why they gave a certain score and this was later transcribed to understand reasons for the ratings.

3.4.3 Participants' Knowledge

The interview was captured on a video camera which was then transcribed. We developed codes (Table 2) to apply to their answers, breaking them into declarative, structural and procedural knowledge types. For each participant, we then calculated a mental model score for each type of knowledge as the sum of the answers where an incorrect answer was valued as -1 and a correct answer was valued as 1. A Mann-Whitney U test was used to compare whether there was a significant difference between the two groups for each knowledge type. To ensure reliability of the coding, we conducted an inter-rater reliability test between two researchers on 20% of the data. Given an acceptable average agreement of 0.63 for the Jaccard Index, the rest of the transcripts were independently coded by the first researcher.

Type	Code	Definition	Example
Declarative	Woodblocks and "snap"	Woodblocks play to help you find the face of the person who is most directly in front of the camera. The woodblocks get higher in pitch the more aligned you are to a face. The "snap" sound that comes with it indicates that the system has found the face.	"The higher the pitch of the woodblocks the closer I was of finding the face".
Declarative	"click"	This sound means the system has detected a person.	"When I heard a sound like this (clicks finger) it means the system has found a person".
Declarative	"knock"	This sound means that the system has found a face.	"If the system makes a knock it means it has seen a face".
Structural	Person ID	The system needs to see the eyes, ears and nose of a person to positively identify him/her, about 4 m away. The system has to be trained to recognize people.	"For the system to identify a face, it needs to see the eyes, nose and mouth".
Structural	Person detect	To detect a person, it needs to see the head (eyes, nose, ears) or the torso (shoulder, chest and arms), about 10 meters away.	"In order to identify a person, it needs to see the head or the torso".
Structural	Debugging	Information regarding why the system is not able to identify someone, about what the cause of the problem is.	"The system is not identifying anyone because it cannot see a face".
Procedural	Actions to be taken	What actions needed or taken to use the system effectively, or how to overcome a problem.	"In order to find someone who is standing up. I would have to look up."

Table 2: Knowledge Type Code Set

3.4.4 Participants' Strategies

During the networking event, the participants' behaviors were captured on a video camera. This afforded us the ability to focus on aspects of the participants' movements (e.g., gaze, pace, bodily comportment). We coded the actions by the participants as shown in Table 3.

Code	Description
Gaze: Straight ahead	Participant were looking straight ahead.
Gaze: Low	Participants gaze was low so they were looking at a downward angle.
Gaze movement: Up and Down/Down and Up	Participants moves their head up and then down or the other way around.
Gaze movement: Left and Right/Right and Left	The participant looks horizontally: left and right or right and left.
Stops walking	The participant was walking but has now stopped.
Walks slowly	The participant walks slowly, often one step at a time and then pausing.
Walks normal pace	The participant is walking at a normal pace.
Finds Recruiter	The participant managed to find the recruiter during the task.
Speaks to Facilitator	The participant speaks to the facilitator. This often happened if they had a question.
Conversation with confederates	The participant was having a conversation with the confederates.

Table 3: Strategies Code Set

We used MAXQDA to apply codes directly to the relevant video section. Each time the participant changed their movement (e.g., was walking but then stopped) the researcher paused the video, created a new timestamp and recorded the behavior. Therefore, each different movement acted as a unit of analysis.

Visualizations were created for the behavioral journey of each participant, using the x-axis for time in minute chunks until finding the recruiter, and each different type of behavior received its own color for clarity. We then sorted these visualizations into groups where the behavioral journeys appeared similar, using an approach adapted from [12].

4 Results

4.1 User Experience

We analyzed whether there were any differences in perceived workload (Figure 3). While all measures were lower in general for the Enhanced group, the overall workload for the Enhanced group (*M=23.10, SD=18.54*) was very similar to the Basic group (*M=23.50, SD=16.61*). A subsequent Mann-Whitney U test confirmed that there were no significant differences between Enhanced and Basic groups (*U = 16.00, p = 0.88)*. Further Mann-Whitney U tests indicated that there were also no significant differences between Enhanced and Basic groups for mental demand (U = 15.00, p = 0.71), physical demand (*U = 13.00, p = 0.29*), temporal demand (*U = 17.50, p = 0.66*), performance (*U = 18.00, p = 0.73*), effort (*U = 13.50, p = 0.31*) and frustration (*U = 18.50, p = 0.74)*. We can also note that the workloads for both groups were generally low which indicates low demand and good satisfaction with the prototype.

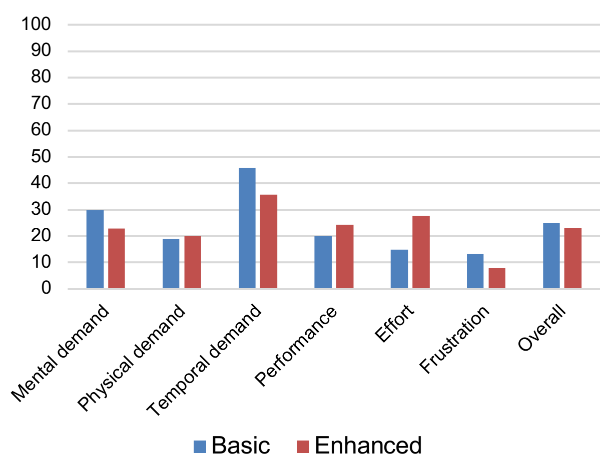

Figure 3: Mean NASA-TLX scores. All scores are low, with no significant difference between Basic and Enhanced groups.

Participants stated a variety of reasons for their ratings. Recall that higher ratings are 'worse'. Mental demand and effort were sometimes rated high because participants had to remember what every sound meant which is very difficult with such limited exposure. Some higher ratings for physical demand were based on weight of the headset. The higher ratings for temporal demand might be due because many participants felt the duration of talking to the recruiter was too short:

> "Really wanted to do it for longer. That's because it's 15 minutes rather than 20 minutes."- User 11_Enhanced

Many participants rated their performance quite low because even though they found the recruiter successfully they had difficulties with misidentifications or figuring out the direction of sounds along the way:

> "[...] I did find her quite successfully but I had difficulty. The problem was that the system recognized <the recruiter> and but apparently there was a person standing in front of me that the system was not trained on so I was a little confused since I thought <the recruiter> was standing in front of me whilst she was not." - User4_Enhanced

We then investigated whether the two kinds of instructions had any bearing on the success of participants in completing the task. When we analyzed system's accuracy we found that both groups had very similar accuracy in identifying the recruiter, M=18.5% accuracy, SD=0.13 for the Basic group and M=15.34%, SD=0.04 for the Enhanced group. Because of the small and imbalanced sample, we did not investigate this through a statistical test. Further, there was not a significant difference between the two groups (*p=0.445*) even though the Basic group (*M=1:13, SD=0.03*) found the recruiter slightly faster than the Enhanced group (*M=2:31, SD=0.09*).

These findings suggest that participants experience of using the system and the workload they have when completing a task does not significantly differ when they do, or do not have, additional knowledge of how the system works. Our results echo Kulesza et al.'s findings [24], where there were no significant differences in demand when participants received additional knowledge about the internal workings of the system compared to those who did not. Therefore, this suggests knowing more about the complex nature of the system does not create a workload burden, contradicting suggestions that 'simpler' explanations should be favored [27].

4.2 Knowledge Gained

The users' mental model of the system was an important aspect in the study as we know from previous research that this shapes how they use the system and predict its actions [23,25]. We therefore turned our attention to the mental model scores that we calculated for each participant, and investigated whether there were any differences in the types of knowledge that participants with basic or enhanced instructions displayed, for each type of knowledge (Figure 4).

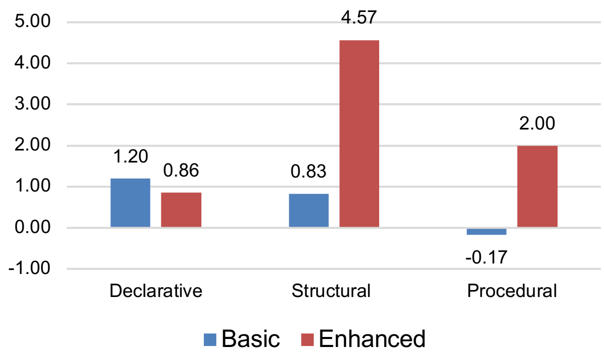

Figure 4. Mean mental model scores. There were no significant differences between Basic and Enhanced groups for declarative knowledge scores but there were significant differences between structural and procedural knowledge scores.

4.2.1 Declarative Knowledge

A Mann-Whitney U test found that there was no significant difference in the declarative knowledge the users were able to recall during the post-task interview (*U = 19.50, p = .84*). This suggests that in both groups participants could recall the meaning of each sound correctly; not surprising as both groups received an explanation of what each sound meant. However, declarative knowledge cannot be guaranteed for everyone. For example:

> "Spin his head -moves his head from left to right- to check what happens with the sound. If the sound stops after a certain degree he can orient the amount of degrees which the person is situated. He can kind of get the dimension of the person."- User10_Basic

Previous research has found users often misunderstand computer vision systems [18]; in this case, the system does not use dimensions of people. This is something that designers of computer vision systems need to be cautious of to avoid losing the user's trust when their expectations are violated [19].

4.2.2 Structural Knowledge

A Mann-Whitney U test found that the mental model scores were significantly different between the Basic and the Enhanced group (*U = 4.00, p = 0.014*). The Enhanced intelligibility instructions delivered structural knowledge of the system, so it makes sense that the Basic intelligibility group scored lower.

Structural knowledge is crucial if the system does not behave as expected. For example, in response to 'What should Tom assume if he the system does not call out a name?':

> "The person is not entered in to the system because if I look at you, I hear the sound but I do not hear anything." - User7_Enhanced
> "that…it may be the case that a person that the system is not trained on is standing between you and the recognized person so I would tell him to watch out for that". -User4_Enhanced

In contrast, the absence of correct structural knowledge of the system often left participants in the Basic intelligibility group to create their own ideas about how the system works:

> "The computer vision system detected the person in front of you and determined that it's <name of researcher>, so it can just be feet, foot, I don't know arm, or hair, anything I guess, as long as it's here."- User6_Basic

Drawing on declarative knowledge, participants in the Basic intelligibility group could have correctly deduced that the system sees a face because they were told that the face identifier knock sound plays "when the camera aligns with a person's face".

However, some participants moved more effortlessly between different knowledge types and knowledge levels, for example:

> "Move his head around slowly listen to the pitch of the woodblocks sound, the higher it gets the closer the face of the person is."- User 4_Enhanced

This participant knew that the computer vision system has to get closer to the persons' face for the pitch to get higher and can articulate a rule for this. He then suggested what behavior should be carried out when the user hears the Woodblocks, moving to the knowledge-based level. Thus, it appears that structural knowledge needs to be explicitly taught, as it is not a given that users will be able to 'reason from first principles' and get from declarative knowledge to structural knowledge easily.

4.2.3 Procedural Knowledge

An example of an incorrect procedural knowledge was when the participant was asked: "If Tom was sitting down and I was standing in front of him about an arm's length away where would Tom have to look in order for the system to recognize me and call out my name?"

> "[points to his face] I would first turn him around [pivots his body around to show a turning around movement] tell him to turn and whenever he hears a little knocking sound he has to stop and walk for a bit. I would start him having his back to the person." – User10_Basic

Although this participant provided correct declarative knowledge – that the system has to see the target's face – the procedural advice he offered is incorrect. It would not make sense to turn your back to the target's face if you are trying to identify who they are. This is an inappropriate way of using the system.

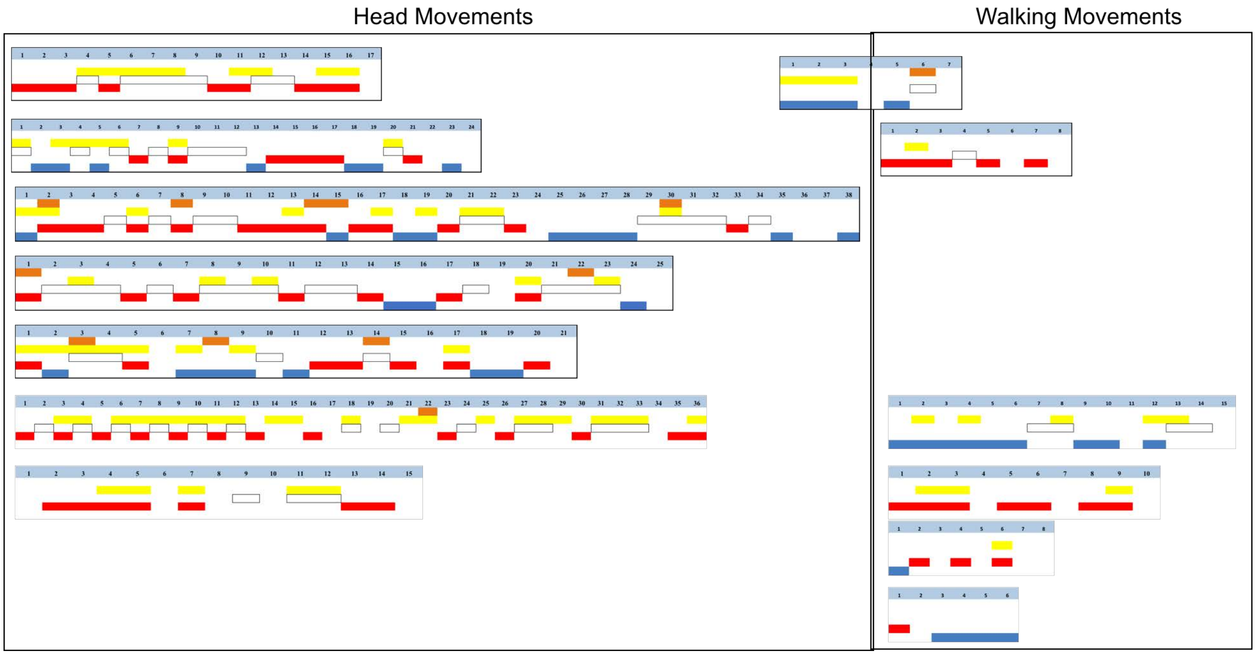

Figure 5: Behavioural Journey For Participants. (Horizontal (yellow) and vertical (orange) head movements, stopping (white), and walking slowly (red) and at a normal pace (blue). Enhanced group participants' journeys are outlined in black.) More participants in the Enhanced group than the Basic group used horizontal head movements to explore their environment, while participants in the Basic group used walking to explore the space.

We conducted a Mann-Whitney U test to determine whether there was significant difference between the Basic intelligibility group and the Enhanced intelligibility group. We found that in the Enhanced group scored significantly higher than those in the Basic group ($U = 6.00$, $p = .035$), showing better procedural knowledge. For example, a correct procedural mental model came from a participant when asked "If Tom was sitting down and I was standing in front of him about an arm's length away where would Tom have to look in order for the system to recognize me and call out my name?":

> "Straight up and above to recognize your face. He would have to judge your height and then towards it."
> – User3_Enhanced

Recall that none of the participants received procedural knowledge of the system as part of their instructions, that is, behaviors that one could take to improve the use of the system. Our finding suggests that users in the Enhanced intelligibility group were able to leverage their structural knowledge about the system and its internal workings to guide them on what behaviors they should adopt when using the system, using a knowledge-based level.

4.3 Strategies Used

While we found very similar user experience across the different groups, we wanted to see whether the participants' different types and levels of knowledge might have had implications for their interactions with the technology.

About half of the participants used horizontal gaze movements, where they scanned their environment from side to side (Figure 5, yellow) to find the recruiter, usually while they stopped walking (Figure 5, white). Five participants using this strategy were in the Enhanced group whereas there were only two participants in the Basic intelligibility group.

Looking side to side is suited to the computer vision system for three reasons: 1. This head movement increases the field of view of the computer vision system, 2. The participants can use the spatialized audio to get better location information, and 3. there is a higher chance of triggering the ID announcements as targets cross the midline of vision. It seems that this was a strategy that was employed deliberately by participants, as referenced in the post-task interview:

> "He should basically try to stand still and look around slowly -scans his head from left to right- because it will mention the name, or it will make the -knocks on the desk- sound if it sees a face and then if it's a recognizable face it will also mention the name of the said person."- User11- Enhanced

Recall that the system also provides help for finding a face through playing "woodblock" sounds that help to align the user to a person's face which helps with identification. To this end we were also interested in whether participants moved their heads vertically, i.e. up and down, during the task (Figure 5, orange). This is a behavior which can increase the effectiveness of the system, especially as the recruiter was sitting down for the first 7 minutes of the task. Only five participants exhibited this behavior, with four of them in the Enhanced intelligibility group. Again, some participants appeared to employ this strategy deliberately, as they mentioned it during the post-task interview:

> "I first looked around a little and then it didn't make any sound or anything then I looked a bit down and it didn't make a sound, then I bend over a little, with my head on the sitting height and then it managed to."- User11- Enhanced

A different common strategy seems was employed by six participants, who explored the environment through walking around slowly (Figure 5, red) or at a normal pace (Figure 5, blue). Five participants using this strategy were in the Basic intelligibility group whereas only one were Enhanced group participants. In contrast to the head movement strategy, this might not lead to better use as announcements might be missed and spatialized audio might not be as useful to these participants.

Taken together the results suggest participants in the Enhanced intelligibility group in comparison to those who received basic instructions were more likely to apply effective movements when using the system, and do so deliberately.

5 Discussion and Conclusions

Our study investigated the intelligibility of a computer vision system for blind users when receiving only basic instructions containing declarative knowledge versus enhanced instructions comprising additional structural knowledge of the system. Our study found that these different types of knowledge played no part in perceived or actual user experience that we were able to measure during the very short time our participants employed this technology. Possibly, with more data, especially over a longer duration of use, this might change, and differences might still be found.

Unsurprisingly, because they were taught this through enhanced instructions, we found that participants in the Enhanced group knew more of how the system worked i.e. they had more structural knowledge. However, these participants also had more procedural knowledge i.e. they knew what to do to make the system work well, and to overcome any obstacles in its use. In the absence of structural and procedural knowledge, users might build up a 'wrong' or incomplete mental model of the prototype which might also reflect in inappropriate use.

We found some evidence that this is indeed the case. We presented results that indicated two distinct strategies in using the prototype, one of which might be more useful in locating and identifying people by blind users. Exploring the environment through scanning either horizontally or vertically is exploiting the prototype's capabilities more effectively, and this strategy was more frequently and deliberately taken by participants receiving Enhanced instructions.

These findings have three clear implications for designing and building computer vision systems for blind users:

- These systems can be successfully employed by blind users with very little training. Almost all of our participants found the right person within a couple of minutes, and the workload of using the prototype was low.
- In addition to basic declarative knowledge, developers of these systems need to give users structural knowledge about how the system works in terms of how features are used in the models and how these relate together to enact the system behavior, in this case, person detection and identification.
- Different strategies of users employing the system are to be expected. However, it appears it is possible for blind users to build better mental models when given more detailed information, translating this into more effective search strategies. It seems that successful use of the system can be 'nudged' through more information, echoing findings from explaining interactive machine learning systems [21,23].

Our study is not without limitations. First, we were only able to observe a very short period of prototype use with limited numbers of participants due to the difficulties in recruiting this user group. Further studies of the prototype system should be extended in scope, and take place in real-world social situations. This would allow us to investigate how users grasp the nuances of the system given the instruction over time, and whether the behavior and mental model of users given basic instructions would improve with further experience of the system, to what was observed for the Enhanced instruction group.

Second, the participants of our study differed in their visual abilities; some participants had some albeit very reduced vision which might have affected how they used the system. Further studies are necessary on how to support blind users especially as there is a wide range of visual abilities within this user group.

Third, we were also stymied by data losses that affected our ability to investigate the effect of intelligibility on the accuracy of the system. However, measuring the accuracy of computer vision systems in their natural use such as ours is difficult. Typically, accuracy is evaluated by precision and recall. However, in a computer vision system that runs continuously and changes views dynamically with head movements, this is difficult to measure as potentially every frame would need manual labeling with ground truth data.

Last, our study only differentiated between basic and enhanced instructions but did not focus on what instructions mattered the most, nor how these instructions could be most effectively delivered. Investigating the mental models of users in detail could provide further insights on more intelligible designs of such computer vision systems, and well as optimizing the method for delivery. For example, users could take part in a demonstration and experience the system first-hand. This would likely further deepen their mental model.

Our study can also inspire future work investigating the intelligibility more generally. We argue that there is a need for exploring the following research gaps :

- Currently, a lot of research is conducted into making systems transparent, without similar research into what would make these systems intelligible to users. Further empirical studies to measure intelligibility of systems are desperately needed to advance our understanding in designing appropriate explicit explanations, or brief instructions, or how to support intelligibility through exploration.

- A much more refined characterization of users' 'understanding' is warranted. In this study, we have attempted to tease apart the correctness of the mental model, and the types and levels of knowledge that a user might have. These differences in 'understanding' might have finely faceted impacts on system use.
- A common effect of explanations and improved mental models that is investigated is trust. However, in many circumstances, appropriate or better system use might be a more desirable outcome. Further investigations of the impacts of intelligibility on human-machine teaming or cooperation between human and computer is needed.

Our study provides a step towards making AI systems intelligible to users, by supporting people who are blind to better use a computer vision system in navigating social situations.

ACKNOWLEDGMENTS

We thank the participants in our study for their time, and VICTA in helping us with recruiting participants. This work was supported by Microsoft Research, providing access to technology and resources to run this study.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 582:1–582:18. DOI:https://doi.org/10.1145/3173574.3174156

[2] A. Adadi and M. Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6, (2018), 52138–52160. DOI:https://doi.org/10.1109/ACCESS.2018.2870052

[3] Vijay Arya, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Q. Vera Liao, Ronny Luss, Aleksandra Mojsilović, Sami Mourad, Pablo Pedemonte, Ramya Raghavendra, John Richards, Prasanna Sattigeri, Karthikeyan Shanmugam, Moninder Singh, Kush R. Varshney, Dennis Wei, and Yunfeng Zhang. 2019. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. *ArXiv190903012 Cs Stat* (September 2019). Retrieved October 7, 2019 from http://arxiv.org/abs/1909.03012

[4] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-aware Systems. *Hum-Comput Interact* 16, 2 (December 2001), 193–212. DOI:https://doi.org/10.1207/S15327051HCI16234_05

[5] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: A Visual Interactive Hybrid Recommender System. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (RecSys '12), 35–42. DOI:https://doi.org/10.1145/2365952.2365964

[6] Rupert R. A. Bourne, Seth R. Flaxman, Tasanee Braithwaite, Maria V. Cicinelli, Aditi Das, Jost B. Jonas, Jill Keeffe, John H. Kempen, Janet Leasher, Hans Limburg, Kovin Naidoo, Konrad Pesudovs, Serge Resnikoff, Alex Silvester, Gretchen A. Stevens, Nina Tahhan, Tien Y. Wong, Hugh R. Taylor, Rupert Bourne, Peter Ackland, Aries Arditi, Yaniv Barkana, Banu Bozkurt, Tasanee Braithwaite, Alain Bron, Donald Budenz, Feng Cai, Robert Casson, Usha Chakravarthy, Jaewan Choi, Maria Vittoria Cicinelli, Nathan Congdon, Reza Dana, Rakhi Dandona, Lalit Dandona, Aditi Das, Iva Dekaris, Monte Del Monte, Jenny Deva, Laura Dreer, Leon Ellwein, Marcela Frazier, Kevin Frick, David Friedman, Joao Furtado, Hua Gao, Gus Gazzard, Ronnie George, Stephen Gichuhi, Victor Gonzalez, Billy Hammond, Mary Elizabeth Hartnett, Minguang He, James Hejtmancik, Flavio Hirai, John Huang, April Ingram, Jonathan Javitt, Jost Jonas, Charlotte Joslin, Jill Keeffe, John Kempen, Moncef Khairallah, Rohit Khanna, Judy Kim, George Lambrou, Van Charles Lansingh, Paolo Lanzetta, Janet Leasher, Jennifer Lim, Hans Limburg, Kaweh Mansouri, Anu Mathew, Alan Morse, Beatriz Munoz, David Musch, Kovin Naidoo, Vinay Nangia, Maria Palaiou, Maurizio Battaglia Parodi, Fernando Yaacov Pena, Konrad Pesudovs, Tunde Peto, Harry Quigley, Murugesan Raju, Pradeep Ramulu, Serge Resnikoff, Alan Robin, Luca Rossetti, Jinan Saaddine, Mya Sandar, Janet Serle, Tueng Shen, Rajesh Shetty, Pamela Sieving, Juan Carlos Silva, Alex Silvester, Rita S. Sitorus, Dwight Stambolian, Gretchen Stevens, Hugh Taylor, Jaime Tejedor, James Tielsch, Miltiadis Tsilimbaris, Jan van Meurs, Rohit Varma, Gianni Virgili, Jimmy Volmink, Ya Xing Wang, Ning-Li Wang, Sheila West, Peter Wiedemann, Tien Wong, Richard Wormald, and Yingfeng Zheng. 2017. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob. Health* 5, 9 (September 2017), e888–e897. DOI:https://doi.org/10.1016/S2214-109X(17)30293-0

[7] Michael Brock and Per Ola Kristensson. 2013. Supporting Blind Navigation Using Depth Sensing and Sonification. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (UbiComp '13 Adjunct), 255–258. DOI:https://doi.org/10.1145/2494091.2494173

[8] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, T. D. Kelley, D. Braines, M. Sensoy, C. J. Willis, and P. Gurram. 2017. Interpretability of deep learning models: A survey of results. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1–6. DOI:https://doi.org/10.1109/UIC-ATC.2017.8397411

[9] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *ArXiv Prepr. ArXiv170208608* (2017).

[10] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *Int. J. Hum.-Comput. Stud.* 58, 6 (June 2003), 697–718. DOI:https://doi.org/10.1016/S1071-5819(03)00038-7

[11] Malin Eiband, Charlotte Anlauff, Tim Ordenewitz, Martin Zürn, and Heinrich Hussmann. 2019. Understanding Algorithms Through Exploration: Supporting Knowledge Acquisition in Primary Tasks. In *Proceedings of Mensch Und Computer 2019* (MuC'19), 127–136. DOI:https://doi.org/10.1145/3340764.3340772

[12] Valentina Grigoreanu, Margaret Burnett, Susan Wiedenbeck, Jill Cao, Kyle Rector, and Irwin Kwan. 2012. End-user Debugging Strategies: A Sensemaking Perspective. *ACM Trans Comput-Hum Interact* 19, 1 (May 2012), 5:1–5:28. DOI:https://doi.org/10.1145/2147783.2147788

[13] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (ed.). North-Holland, 139–183. Retrieved November 6, 2013 from http://www.sciencedirect.com/science/article/pii/S0166411508623869

[14] Diane Warner Hasling, William J. Clancey, and Glenn Rennels. 1984. Strategic explanations for a diagnostic consultation system. *Int. J. Man-Mach. Stud.* 20, 1 (January 1984), 3–19. DOI:https://doi.org/10.1016/S0020-7373(84)80003-6

[15] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241–250. DOI:https://doi.org/10.1145/358916.358995

[16] J. Karreman, N. Ummelen, and M. Steehouder. 2005. Procedural and declarative information in user instructions: what we do and don't know about these information types. In *IPCC 2005. Proceedings. International Professional Communication Conference, 2005.*, 328–333. DOI:https://doi.org/10.1109/IPCC.2005.1494193

[17] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, and Rory Sayres. 2018. Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In *International Conference on Machine Learning*, 2668–2677. Retrieved December 11, 2018 from http://proceedings.mlr.press/v80/kim18d.html

[18] Jacob Kittley-Davies, Ahmed Alqaraawi, Rayoung Yang, Enrico Costanza, Alex Rogers, and Sebastian Stein. 2019. Evaluating the Effect of Feedback from Different Computer Vision Processing Stages: A Comparative Lab Study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 43:1–43:12. DOI:https://doi.org/10.1145/3290605.3300273

[19] René F. Kizilcec. 2016. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 2390–2395. DOI:https://doi.org/10.1145/2858036.2858402

[20] Peter Kontschieder, Jonas F. Dorn, Cecily Morrison, Robert Corish, Darko Zikic, Abigail Sellen, Marcus D'Souza, Christian P. Kamm, Jessica

Burggraaff, Prejaas Tewarie, Thomas Vogel, Michela Azzarito, Ben Glocker, Peter Chin, Frank Dahlke, Chris Polman, Ludwig Kappos, Bernard Uitdehaag, and Antonio Criminisi. 2014. Quantifying Progression of Multiple Sclerosis via Classification of Depth Videos. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014* (Lecture Notes in Computer Science), 429–437.

[21] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W. Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, 3–10. DOI:https://doi.org/10.1109/VLHCC.2013.6645235

[22] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (IUI '15), 126–137. DOI:https://doi.org/10.1145/2678025.2701399

[23] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (CHI '12), 1–10. DOI:https://doi.org/10.1145/2207676.2207678

[24] Todd Kulesza, Weng-Keen Wong, Simone Stumpf, Stephen Perona, Rachel White, Margaret M. Burnett, Ian Oberst, and Andrew J. Ko. 2009. Fixing the Program My Computer Learned: Barriers for End Users, Challenges for the Machine. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (IUI '09), 187–196. DOI:https://doi.org/10.1145/1502650.1502678

[25] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. *Why and why not* explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09), 2119–2128. DOI:https://doi.org/10.1145/1518701.1519023

[26] Zachary C. Lipton. 2016. The Mythos of Model Interpretability. *ArXiv160603490 Cs Stat* (June 2016). Retrieved from http://arxiv.org/abs/1606.03490

[27] Tania Lombrozo. 2006. The structure and function of explanations. *Trends Cogn. Sci.* 10, 10 (October 2006), 464–470. DOI:https://doi.org/10.1016/j.tics.2006.08.004

[28] Steve Mann, Jason Huang, Ryan Janzen, Raymond Lo, Valmiki Rampersad, Alexander Chen, and Taqveer Doha. 2011. Blind Navigation with a Wearable Range Camera and Vibrotactile Helmet. In *Proceedings of the 19th ACM International Conference on Multimedia* (MM '11), 1325–1328. DOI:https://doi.org/10.1145/2072298.2072005

[29] Tim Miller. 2017. Explanation in Artificial Intelligence: Insights from the Social Sciences. *ArXiv170607269 Cs* (June 2017). Retrieved from http://arxiv.org/abs/1706.07269

[30] Cecily Morrison, Kit Huckvale, Bob Corish, Richard Banks, Martin Grayson, Jonas Dorn, Abigail Sellen, and Sân Lindley. 2018. Visualizing Ubiquitously Sensed Measures of Motor Ability in Multiple Sclerosis: Reflections on Communicating Machine Learning in Practice. *ACM Trans Interact Intell Syst* 8, 2 (July 2018), 12:1–12:28. DOI:https://doi.org/10.1145/3181670

[31] Don Norman. 1983. Some observations on mental models. In *Mental Models*. Lawrence Erlbaum Associates, Hillsdale,New Jersey,US.

[32] Don Norman. 1989. *The Design of Everyday Things*. Currency-Doubleday, New York.

[33] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors* 39, 2 (June 1997), 230–253. DOI:https://doi.org/10.1518/001872097778543886

[34] Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2018. Manipulating and Measuring Model Interpretability. *ArXiv180207810 Cs* (February 2018). Retrieved from http://arxiv.org/abs/1802.07810

[35] J. Rasmussen. 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Syst. Man Cybern.* SMC-13, 3 (May 1983), 257–266. DOI:https://doi.org/10.1109/TSMC.1983.6313160

[36] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16), 1135–1144. DOI:https://doi.org/10.1145/2939672.2939778

[37] Larry R. Squire. 2004. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 3 (November 2004), 171–177. DOI:https://doi.org/10.1016/j.nlm.2004.06.005

[38] Lee Stearns and Anja Thieme. 2018. Automated Person Detection in Dynamic Scenes to Assist People with Vision Impairments: An Initial Investigation. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '18), 391–394. DOI:https://doi.org/10.1145/3234695.3241017

[39] Simone Stumpf, Vidya Rajaram, Lida Li, Weng-Keen Wong, Margaret Burnett, Thomas Dietterich, Erin Sullivan, and Jonathan Herlocker. 2009. Interacting meaningfully with machine learning systems: Three experiments. *Int J Hum-Comput Stud* 67, 8 (2009), 639–662.

[40] 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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 203:1–203:14. DOI:https://doi.org/10.1145/3173574.3173777

[41] Nava Tintarev and Judith Masthoff. 2007. Effective explanations of recommendations: user-centered design. In *Proceedings of the 2007 ACM conference on Recommender systems*, 153–156. DOI:https://doi.org/10.1145/1297231.1297259

[42] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 601:1–601:15. DOI:https://doi.org/10.1145/3290605.3300831

[43] Daniel S. Weld and Gagan Bansal. 2018. The Challenge of Crafting Intelligible Intelligence. *ArXiv180304263 Cs* (March 2018). Retrieved October 7, 2019 from http://arxiv.org/abs/1803.04263

[44] Gesa Wiegand, Matthias Schmidmaier, Thomas Weber, Yuanting Liu, and Heinrich Hussmann. 2019. I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19), LBW0163:1–LBW0163:6. DOI:https://doi.org/10.1145/3290607.3312817

[45] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 215:1–215:14. DOI:https://doi.org/10.1145/3173574.3173789