SNaSI: Social Navigation through Subtle Interactions with an AI agent

Rébecca Kleinberger[1], Joshua Huburn[2], Martin Grayson[3] , Cecily Morrison[4]

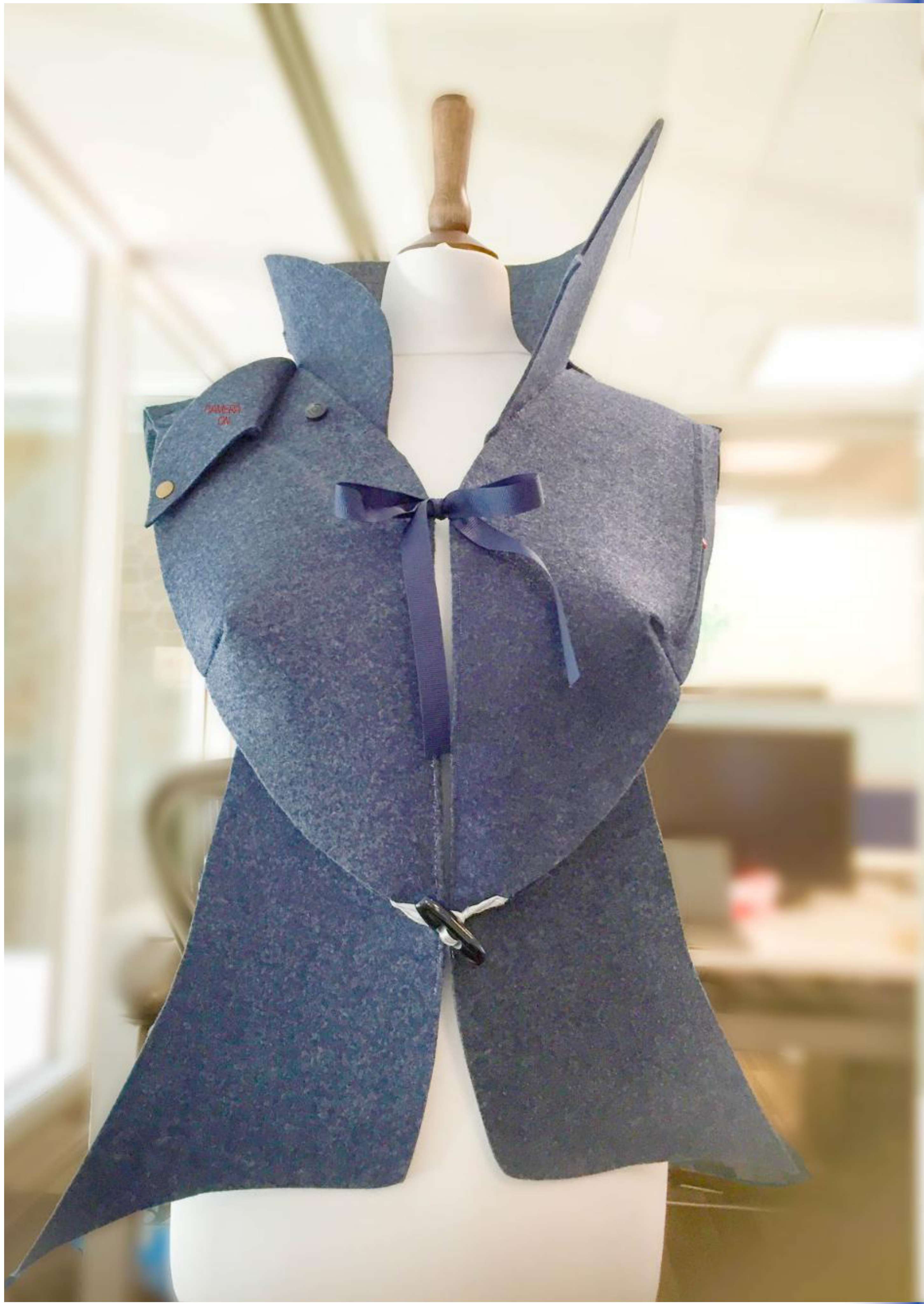
[1] MIT Media Lab, Cambridge, US
& Microsoft Research,
HXD, Cambridge, UK
rebklein@media.mit.edu

[2, 3, 4] Microsoft Research
HXD, Cambridge, UK,
jhuburn@gmail.com
mgrayson@microsoft.com
cecilym@microsoft.com

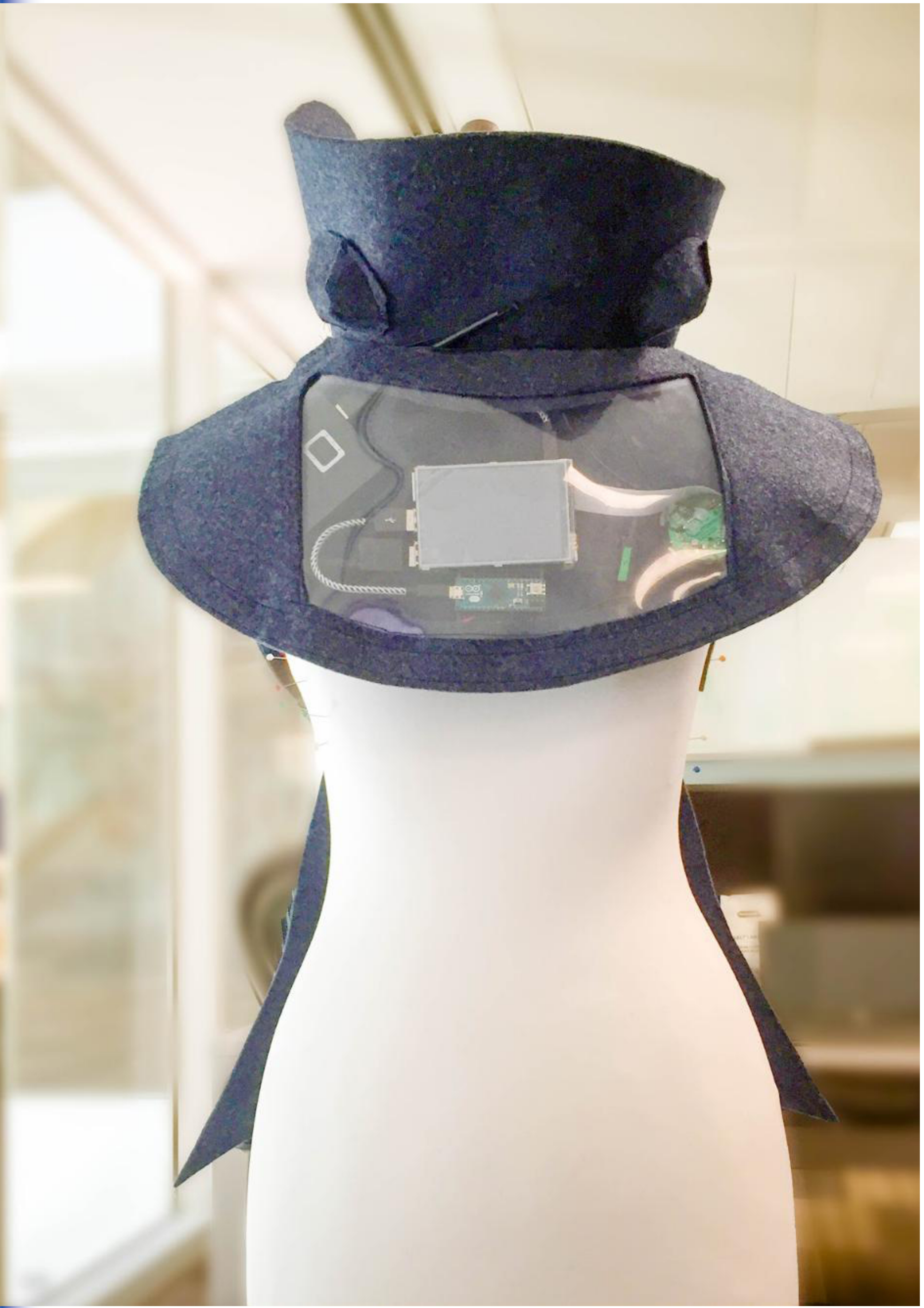
Keywords: Vision Impaired, AI, Wearable Fashion, Assistive Device

Abstract: Technology advances have set the stage for intelligent visual agents, with many initial applications being created for people who are blind or have low vision. While most focus on spatial navigation, recent literature suggests that supporting social navigation could be particularly powerful by providing appropriate cues that allow blind and low vision people to enter into and sustain social interaction. › A particularly poignant design challenge to enable social navigation is managing agent interaction in a way that augments rather than disturbs social interaction. Usage of existing agent-like technologies have surfaced some of the difficulties in this regard. In particular, it is difficult to talk to a person when an agent is speaking to them. It is also difficult to speak with someone fiddling with a device to manipulate their agent. In this paper we present SNaSI, a wearable designed to provoke the thinking process around how we support social navigation through subtle interaction. Specifically, we are interested to generate thinking about the triangular relationship between a blind user, an communication partner and the system containing an AI agent. We explore how notions of subtlety, but not invisibility, can enable this triadic relationship. SNaSI builds upon previous research on sensory substitution and the work of Bach-y-Rita (Bach-y-Rita 2003) but explores those ideas in the form of a social instrument.

Method&
Critique



CAMERA
ON



Introduction

The design of applications that utilize artificial intelligence is receiving substantial research and industry attention. Improvements in computer vision perception and speech interfaces has set the stage to create (artificial) intelligent agents with (computer) vision. While explorations in this area are intended for the mainstream population (Luger & Sellen 2016), a large number of applications are being developed for people who are blind or have low vision (Wu et al. 2017; Kacorri et al. 2017), or have been adopted as such (Pradhan et al. 2018).

Most systems designed for people who are blind or have low vision have focused on spatial navigation: indoors (Sato et al. 2017; Flores & Manduchi 2018), outdoors (Campbell et al. 2014; Fiannaca et al. 2014), and most recently, virtually (Albouys-Perrois et al. 2018; Zhao, Bennett, et al. 2018). However, recent literature suggests that blind and low vision people across cultures would appreciate richer cues when making sense of their social surroundings while staying in line with social rules (Thieme et al. 2018; Panchanathan, S. Chakraborty & McDaniel 2016; Morrison et al. 2017).

Social navigation, in contrast to spatial navigation, can be thought of as the ability to enter into and sustain social interaction. Research has begun to tackle technical aspects of recognizing relevant information for social navigation: identifying people captured in a photo (Schroff et al. 2015); gaze interpretation (Qiu et al. 2016); and presentation of facial cues (Bala et al. 2014; Murray et al. 2016). However, little has been written on how these technological advances could be presented to a blind or low vision user in an agent experience that supports social navigation.

A particularly poignant design challenge to enable social navigation is managing agent interaction in a way that augments rather than disturbs social interaction. In this paper we present SNaSI, a wearable designed to support social navigation through subtle interaction. Specifically, we are interested to provoke thinking about the triangular relationship between a blind user, an communication partner and the system containing an AI agent. We explore how notions of subtlety, but not invisibility, can enable this triadic relationship. SNaSI build upon previous research on sensory substitution and the work of Bach-y-Rita (Bach-y-Rita 2003) but explores those ides in the form of an social instrument

Designing SNaSI

Design for Augmentation

A key design tenant is that an intelligent agent for social navigation should augment human capability through providing cues about the social environment. The cues and modality for transmission are chosen to enrich the human interaction without disturbing the connection. Specifically, the agent is not intended as a utilitarian replacement for vision, providing information a sighted person might otherwise see. Rather, we begin with how we can enrich the well developed sense-making capabilities of people who are blind or low vision with relevant social information (Thieme et al. 2018). We envision the agent as another information source that the user can manipulate. A user may prefer more or less information or have a wider (e.g. room) or specific (e.g. person) focus. Those preferences may be situational or person-specific depending on alternative abilities and personality.

Subtlety

Imagining the agent as an augmentation of sense-making rather than a replacement of vision, forefronts the human-human interaction. Design choices need to forefront that human connection too, yet we cannot consider only the human interaction. The interaction with the human conversational partner remains a dialogue but with three voices – a triadic relationship (as per Figure 2). To facilitate the interaction between people, the communication partner must be aware of what is happening between agent and user without being distracted by it or impacting the privacy of the user. For example, if the agent is speaking to the user, the communication partner should not be speaking too, but the system should not reveal the information requested by the user. As such, an agent experience requires subtlety but not invisibility in design.

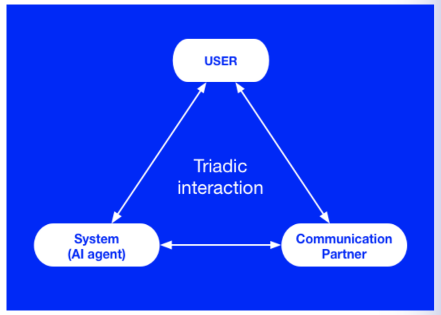


Figure 2. SNaSI is designed to allow a triadic interaction between a user, a communication partner and the wearable system containing the AI agent. This triadic interaction is still a dialogue as the objective of the system is designed to support the human connection between the user and their partner.

Design Choices

SNaSI is a fashionable wearable for social navigation that employs subtle, but not invisible ways to interact with a visual agent. It contains a lapel camera, spatialized audio through two speakers in the collar, a zipper potentiometer and button microphone for user direction of the agent. Processing power is located in a visible or coverable pouch behind the collar. We discuss the particular choices below.

> “we made the choice of creating a garment that is a statement as itself”

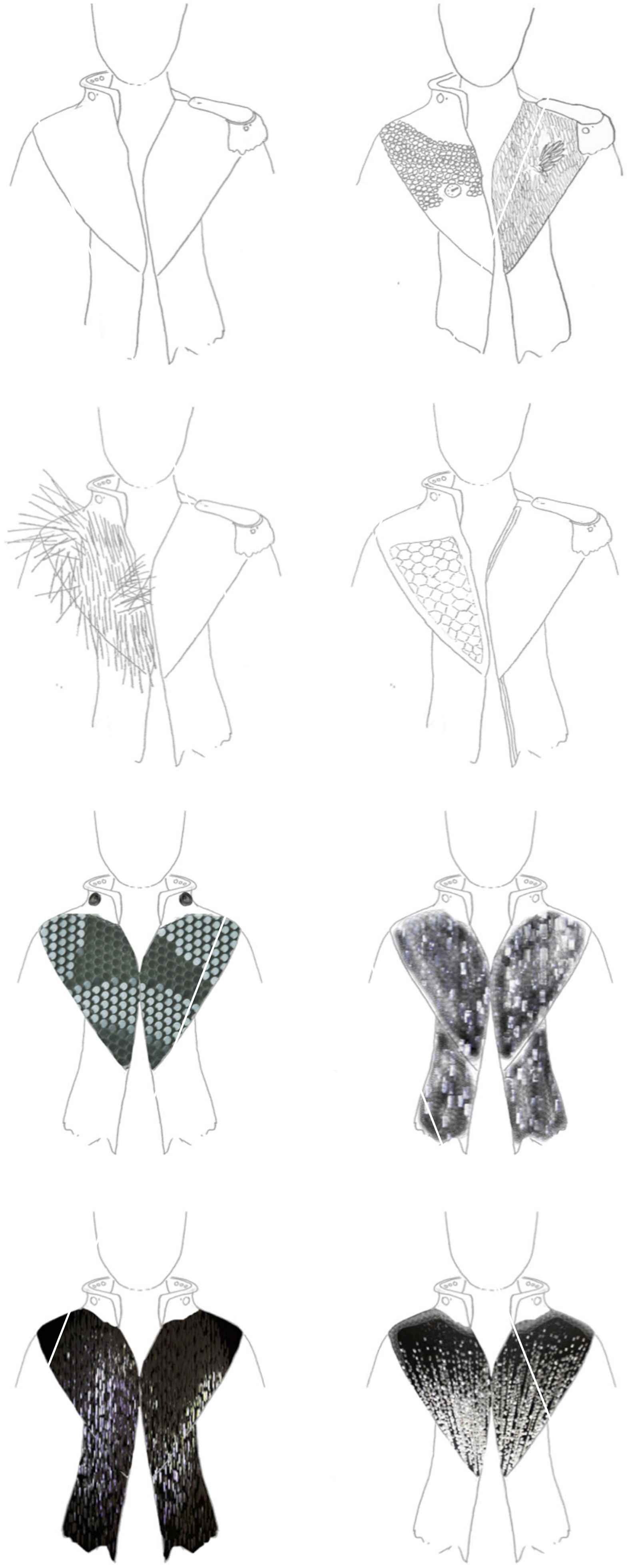


Figure 3. Early sketches of the garment. Through sketching we explored the possible roles of the collar as well as an epaulette to place the camera at a location that would gather the needed visual information without becoming too prominant and impeding the interaction. The shape of the collar continued evolving to contain the 3D audio delivery system. The use of asymmetry allowed the creation of a unique aesthetic.

Form factor

We considered several potential form factors. Existing form factors are often glasses (Ye et al. 2014, Rabia et al. 2014). However, such small devices are not appropriate for computationally intensive algorithms (e.g. identity recognition). Virtual Reality devices are an exception (e.g. Hololens), but appearance would very likely impact human-human interaction. Non-worn devices, such as phones (Zhao, Wu, et al. 2018) or an augmented cane (Gallo et al. 2010), can be difficult to manipulate such that the camera is always in the correct place without substantial cognitive effort and distraction from the social interaction.

Existing wearable systems are often associated with assistive systems and generally have the look of it – bulky and hospital-like equipment – or are invisible and discreet (Jafri et al. 2014). With SNaSi, we made the choice of creating a garment that is a statement as itself. It is fashion while being ergonomic and obviously visible. As a choice of the user, the technology can be either hidden under a protective layer of felt or revealed through a clear window in the back. The garment is to be worn on top of clothing and contains several details that are thought to be aesthetic as much as useful.

The extension of the garment in the front of the body can also serve as a mechanism for interaction with both user and communication partner. The user can feel through actuators signals on the body. The partner can observe visual signals that give indications to agent activity without revealing content. For example, it would be possible to display when the agent is providing information without saying what that information is. Such information enables the communication partner to appropriately time their actions. These features were not implemented in this version, but influenced form factor choice.

The material used for the garment is a 3mm synthetic dark blue felt that gives the advantage of being easily moulded into shape while remaining soft and flexible. The material is laser cuttable and does not require hems. The collar is a tall flared Elizabethan asymmetrical collar that contains the speakers and microphone, and would allow placement of a head rotation sensor in the future. The left shoulder contains a camera easily covered with a fabric shutter and a series of augmented zippers. Figure 3 represents early inspirational sketches of the design, Figure 4 shows the garment worn by a subject and figure 7 shows the front and back view of the final form factor with its hardware components highlighted.

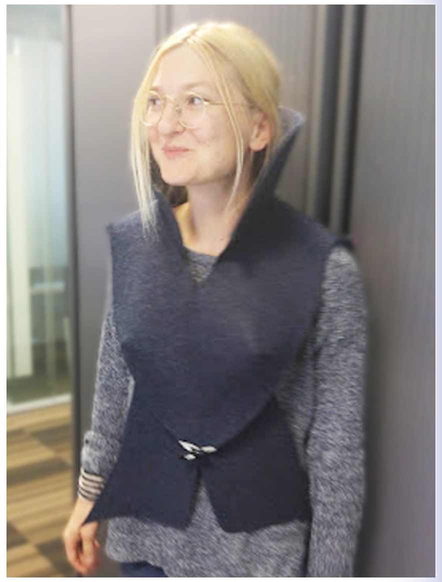


Figure 4. SNaSI garment.

Camera choice

We considered many design options for the camera. Our first idea was to use a pair of cameras on both sides to increase field of view and allow for 3D reconstruction in the central region. The idea was to imitate aspects of the human vision with a center visual field for distance and volume reconstruction and a wide peripheral field mainly used for motion detection. Because of the fluidity and flexibility of the garment, it became difficult to keep the relative position of the camera fixed. Therefore, we made the choice of using only one HD camera on the lapel with a fish-eye lens.

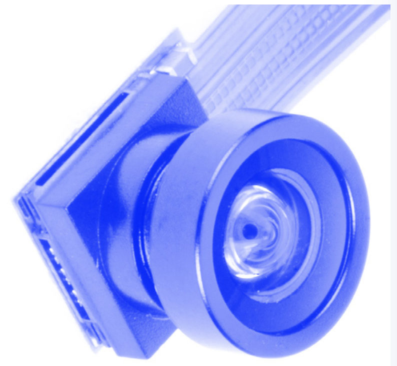


Figure 5. Camera and fisheye lens are placed behind a manually movable shutter that reveals the camera while also displaying a embroided message reading “camera on”.

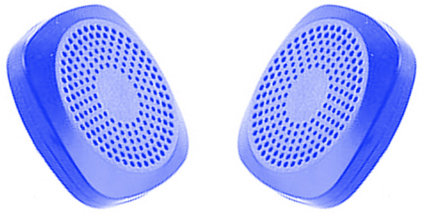


Figure 6. A pair of small speakers placed inside the collar on the left and right. By panning and manipulating the concentration of the sound, the audio gives clue about localization of people in the room.

This specific location allowed for a good balance between three parameters: 1) being frontal and at a good height, it offers a good viewing angle that envelops most of visual field of a sighted user; 2) being located on the lapel, it remains quite discreet compared to being placed directly on the head, glasses, or at head height of the user, which could distract sighted peers accustomed to looking toward the face/eyes of their communication partner; and 3) being placed in a visible location, it respects the privacy of communiation partners by making it obvious when it is in use (a red LED flashes when the camera is on).

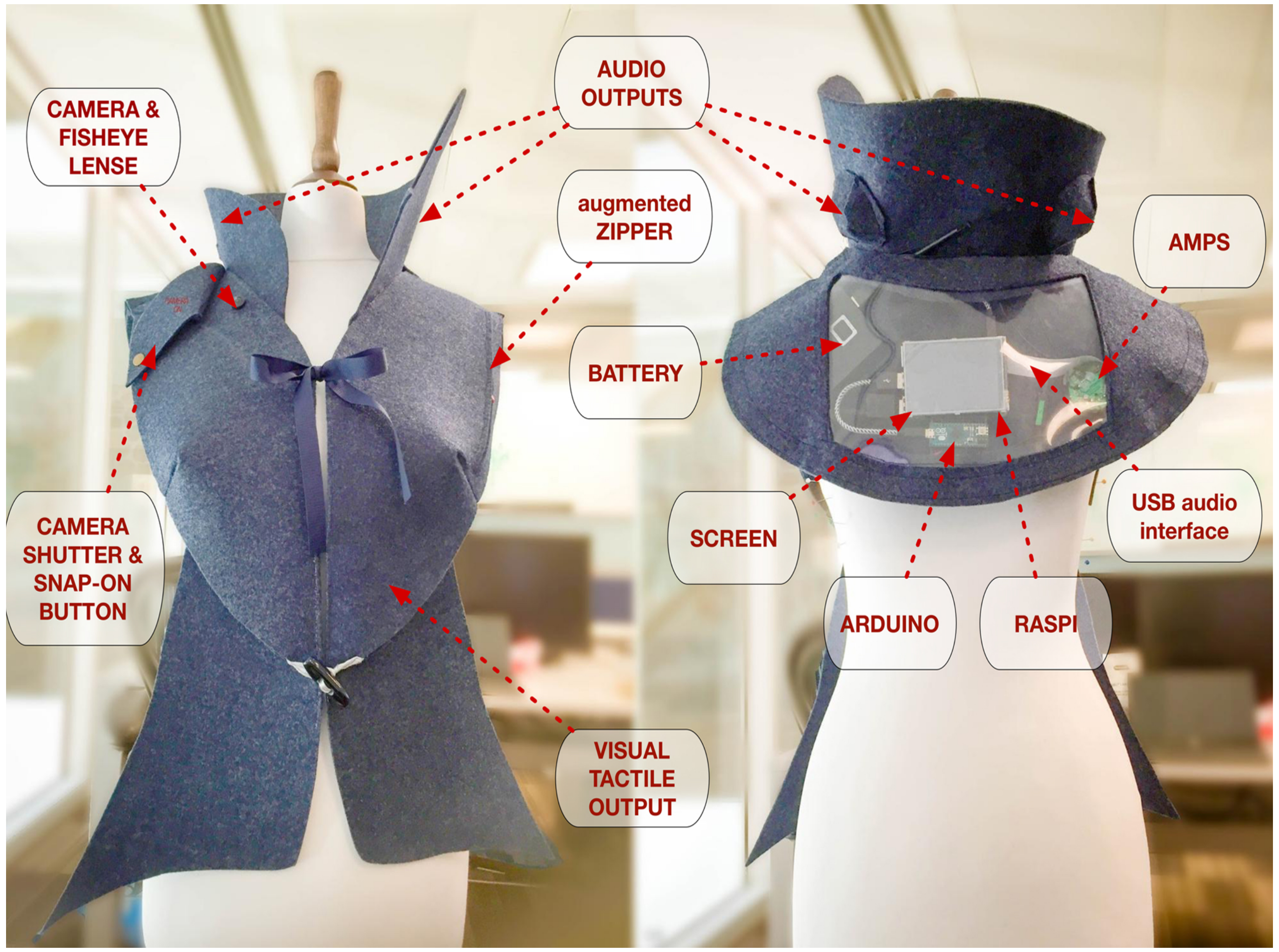
The camera is turned on and off by a small fabric shutter that can cover or uncover the lens. The shutter contains a snap fastener that can snap in two different positions, either hiding the camera or unveiling it. The snap itself is conductive and used as an on/off button to turn the camera on only when the shutter is snapped open. In addition, when snapped open, the fabric shutter displays a camera with an associated red LED light. Taking lessons from research about cameras (Sellen et al. 2007), it was important for the physical appearance to be very clear when an agent was in use as a first step to establish common ground with a communication partner.

Speakers

We made the non obvious choice of using speakers instead of headphones or a bone conduction headset. Headphones block the ears, disturbing conversation and bone conduction headsets have poor fidelity in the low frequencies. Instead, our choice to use speakers was inspired by the way blind people use conversational agent technology (Ye et al. 2014). The use of two speakers and lateral panning provide an approximation of localized audio.

The low volume and rapid pace of messages enable the user to manage and respond to cues while indicating to the conversation part-

Figure 7. The SNaSI garment and its hardware components. The input sensors are a fisheye camera and zipper potentiometers, the main output is the 3D audio speaker system (amplifiers and voice coils) contained in the collar. The longer parts on the front were originally designed to possibly support a visual-tactile display but it not present in the current prototype.



ner that an interaction is happening without providing specifics . The audio cues are personalisable on three axes: abstraction, focus and interpretation. The first axis represents the level of abstraction in which the user wants to receive information. When this level is low, the system provides the information verbally using spoken words. When this level is high, the system provides information through an abstract soundscape whose parameters (granularity, speed, pitch, etc.) represent learned aspects of the interaction. The second axis represents the focus of the situation. The challenges faced by blind and low vision people are very different when in one-to-one interaction compared to a group setting. The third axis represents the level of interpretation wanted. On one extreme of the spectrum, the system would use machine learning and AI to interpret the behaviors of the communication partner and deliver messages such as "your communication partner is not currently listening" or "your communication partner is upset". On the other extreme, the system would feed more subjective data such as a note changing pitch to represent the direction where the communication partner is looking.

Zipper Potentiometer

A series of zipper potentiometers was created to enable a discreet way for the user to shift the scope of information that they receive. Inspired by the field of ambient device design (Wisneski et al. 1998), we wanted an interaction which would be considered "normal" for textiles so that its usage would not distract from the social interaction. We were able to achieve this by adding three custom made zippotentiometers (zipper potentiometers) into the garment. We used conductive thread sewn into regular zippers and connected those to the Arduino to sense the continuous position of the zippers. The zipper becomes a way to talk to the system. Most of us go through our days with cognitive overload. For people who are blind or low vision, some activities and interactions can take more of their energy. Being sensitive about giving just the right amount of information at the right moment is very important for good usability. By simply zipping up or down zippers via a very well known tactile interaction, the user controls the level of focus, abstraction and interpretation of the message delivered by the AI agent.

Processor

The processing is done on an Arduino board which sends the data to a Raspberry Pi embedded linux computer. The Pi receives the frames from the camera as well as the Arduino signals and processes them with Arduino, Python and Pure Data. We choose to use Arduino and Pi because they are widely used in the tech community, inexpensive and offer high potential for personalization. We can, indeed, imagine that a user experienced in digital technologies would like to personalize their own system.
As indicated in Figure 4, the Arduino board, Pi board, battery, amps and USB audio interfaces are located in the back collar of the garment. In addition, we added a small display attached to the Pi that enables display of the screen of the Linux machine run on the Pi. This is mainly used for programming and debugging the embedded system. We could imagine this screen also being used as an interface in the future to allow a curious sighted peer to understand what is analyzed or recognized from the captured field, or for a sighted aid to help parameterize the system. The garment offers two options of covers for back collar that Velcro on to the hidden processor compartment. One cover is in thick felt hiding the processor or one cover is made in clear plastic reveling the system and the screen.

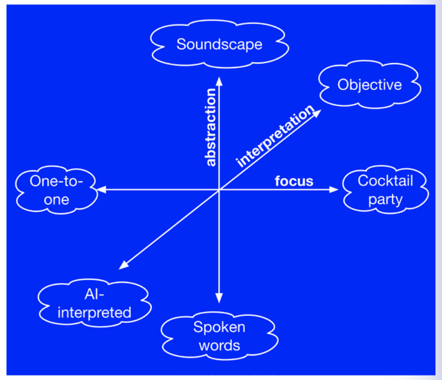
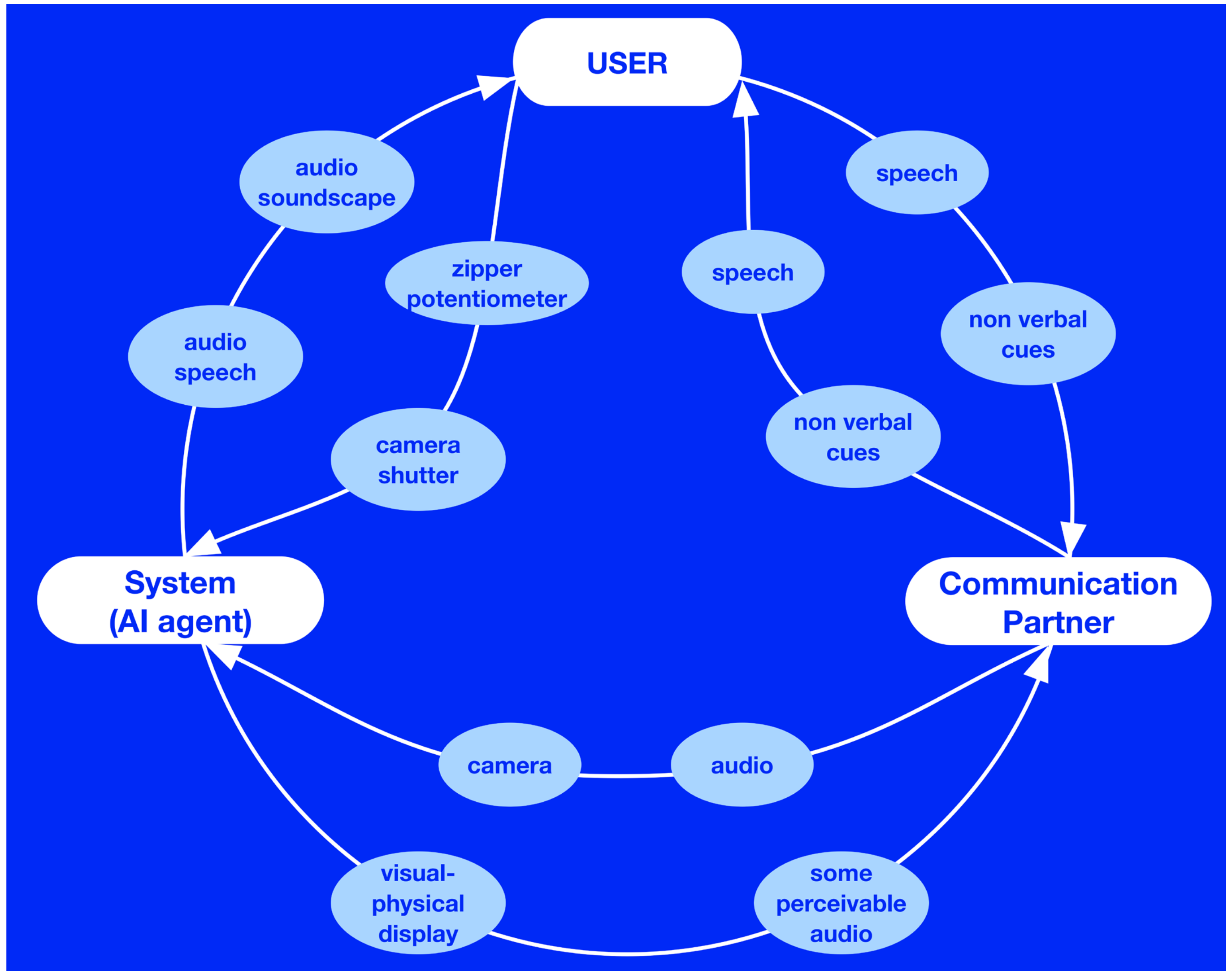


Figure 8. Map of parameters of the system organised on the 3 axes: abstraction, focus and interpretation. Each parameter is controlled by its own associated zipper

Figure 10 (right). Triangular interaction between the user, the system and the communication partner. In this triangle, each actor connects to the two others bidirectionally. The AI gets information from the user (parameters, on/off) and provides them with audio information. The AI also gathers information about the partner through their speech and non verbal cues collected by the camera. The user receives two types of audio cues from the AI agent, some as spoken words and some as audio soundscape. They also receives verbal and non-verbal vocal cues from the partner. The partner can perceive some unintelligible audio coming from the AI that let them know that some information is being transmitted to the user. In a next version, the wearable could also display some information in the form of a visual physical display for both the user and the partner to understand.

Figure 9. By sewing conductive thread along the two sides of a regular metal zipper we created a zippotentiometer capable of sensing the precise percentage of opening/closing of the zipper. When added to the garment, it allows the user to have continuous control of the system.



User Experience

As the technology in this space is rapidly developing, we went with the assumption that our system would have the following computer vision signals. The image frames received from the camera are analyzed by a computer vision module (CV) that focuses on detecting people, their identity, position, and posture. If several people are present, the CV module can estimate their position in the room compared to the orientation of the user. The CV module coupled with the AI agent detects if a person has been seen before and is in the database as well as give an estimate of their age and gender. If the people are closer than a certain distance, the CV module can detect their gaze direction, head, and body orientation as well as facial expression. When combined with the AI agent knowledge, the system will detect if the person seems focussed, bored, happy or unhappy. In addition, by combining past information about the conversational partner, the system can infer for example if the person is paying attention to the user or has been fidgety and not physically oriented toward the user.

To offer the reader with a description of the auditory experience, we present two examples of situations and how changing zipper settings will change the message received by the user: (right margin)

- **Spoken words - Subjective - One-to-One:**
"Tom is not paying attention."
- **Spoken words - Objective - One-to-One**:
"Tom turned his head 30° down and is oriented 15° away to the left."

- **Spoken words - Subjective - Cocktail Party:**
"man - 35yo - 45° left - 6ft away; woman - 56yo - 20° right - 11ft away; woman - 12yo - 50° right - 5ft away".
- **Soundscape - Subjective - Cocktail Party:**
Musical sounds play panning from left to right as the CV modules scan the room and for each person detected, the sound will change with the gender (harp vs. piano note), age (note pitch high to low), distance (loudness) of the person detected.

In these two scenarios, the audio output is delivered by the AI agent in a subtle way. It brings extra information to the user to support them engaging with and managing the conversation. We wanted such information to support blind people in their desire to initiate conversations rather than just react to others.

Evaluation & Future Work

We have been informally engaging with a range of people with visual impairments to inspire the thinking process. Discussions with blind individuals for example, steered us to focus on the fashion of the device as a key priority.

Blind collaborators were also keen on sharing their thoughts on which situations such a design could help improve. For instance, one person thought it could be handy for silent greetings. When a friend or colleague passes by, looks at you, smile, and nods silently. “In that case, the design could socially be of huge personal value!”. Through the process of iteration and design, we were able to bring upon questions, challenges, and insights from our blind collaborators. It was very enriching for defining the design space and segmenting the problem into our three axes in our map of parameters in order for the AI to cover a wide range of social navigation situations.

An interesting next step would be to run an evaluation and see how our thinking plays out in actual conversations. This is a challenge given the fragility of the technology, but an important step to further validate the design thinking produced in the creation of this wearable.

Design Reflections

The design of the SNaSI garment is based on subtlety but not invisibility to enable a triadic relationship that ultimately enriches the dialogue between a blind user and a communication partner. By developing this wearable we designed a physical language that is both understandable by the machine and by the human communication partner but is interpreted differently by each party. The taxonomy enabled by the design works in three steps:

First, when opening the fabric shutter and revealing the camera, the wearable expresses clearly that a third entity has joined the conversation. Done in a casual way, this allows the communication partner to enter a different frame of mind. The social acceptability of the system is enabled not only by the context of being used by the blind person, but also by the way the wearable is presented in a casual though obvious way to the communication partner. The partner needs to know that there is an other entity, that something is happening in the background, but only so that they might monitor and understand the attention of the user when conversing. Second, our wearable has different ways to provide information to the user while providing the communication partner with enough cues to be able to monitor the user’s attention. When information is transmitted through the speakers, even though the partner does not understand the message, they know that something is being transmitted. This might allow them to temporally modulate their own behavior and reduce the information load. Symmetrically, the system knows not to transmit information while the partner is talking to avoid interrupting the human interaction. When the partner is speaking, social information is often accessible to the user (location, level of engagement, gender, age range, emotional state, etc.) Thus the system avoids redundancy.

Third, the system changes behaviors, creating interaction opportunities that were not previously possible. The additional information provided to the user allows them to gain more control of the situation. For instance, the user might now be able to locate even silent people in a room and actively engage in conversation with them without having to wait for the other person to initiate. Or if the user knows that someone is approaching from behind, they can turn around and engage in a new conversation. This ultimately helps create and reinforce a common ground between the blind user and the conversational partner.

In conclusion, the focus on subtlety, but not invisibility achieves an agent that augments but does not disturb social interaction.

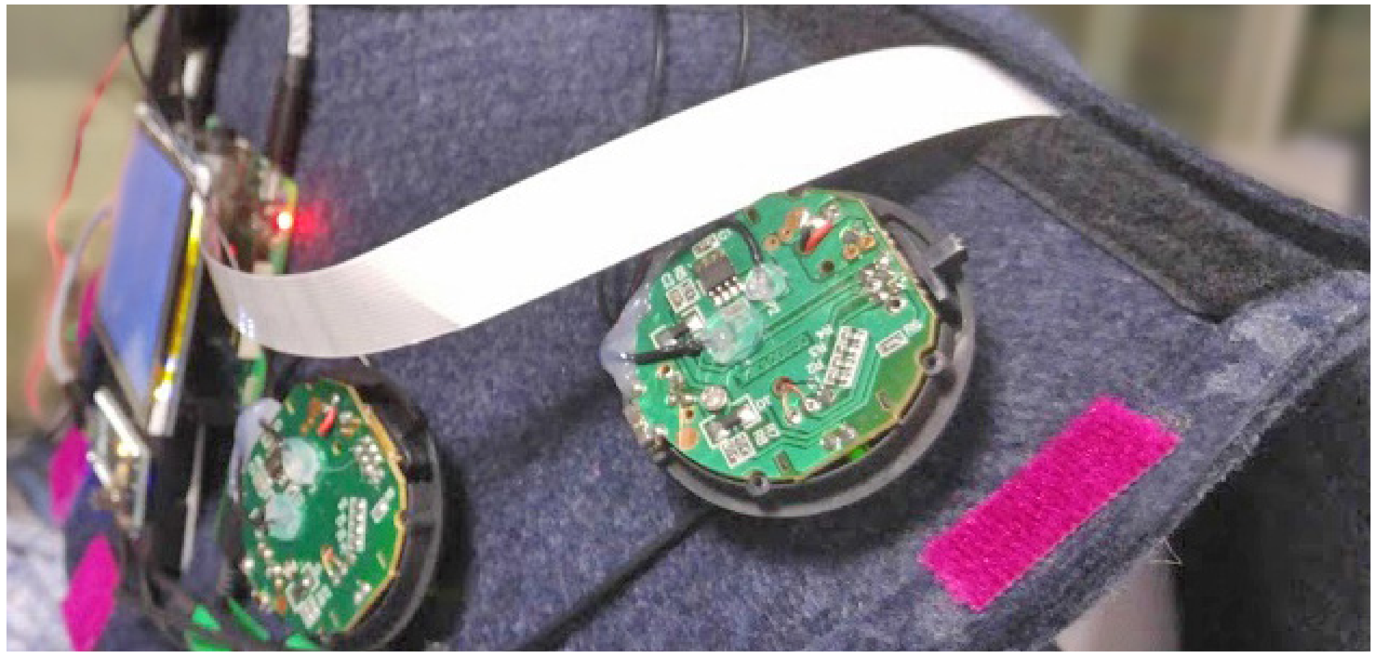


Figure 12. Hardware revealed on SNaSI shoulder piece with removed cover

References

Albouys-Perrois, J. et al., 2018. Towards a multisensory augmented reality map for blind and low vision people: A participatory design approach. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 629.

Bala, S., McDaniel, T. & Panchanathan, S., 2014. Visual-to-tactile mapping of facial movements for enriched social interactions. In Proceedings of the International Symposium on Haptic, Audio and Visual Environments and Games (HAVE). pp. 82–87.

Bach-y-Rita, P. and Kercel, S.W., 2003. Sensory substitution and the human–machine interface. Trends in cognitive sciences, 7(12), pp.541-546

Campbell, M. et al., 2014. Where’s my bus stop?: supporting independence of blind transit riders with StopInfo. In Proceedings of the 16th international ACM SIGACCESS Conference on Computers & Accessibility. pp. 11–18.

Fiannaca, A., Apostolopoulous, I. & Folmer, E., 2014. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. In Proceedings of the 16th international ACM SIGACCESS conference on Computers & Accessibility. pp. 19–26.

Flores, G. & Manduchi, R., 2018. Easy Return: An App for Indoor Backtracking Assistance. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 17.

Gallo, S., Chapuis, D., Santos-Carreras, L., Kim, Y., Retornaz, P., Bleuler, H. and Gassert, R., 2010, September. Augmented white cane with multimodal haptic feedback. In Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference pp. 149-155 .

Jafri, R. and Ali, S.A., 2014, September. Exploring the potential of eyewear-based wearable display devices for use by the visually impaired. In User Science and Engineering (i-USEr), 2014 3rd International Conference pp. 119-124.

Kacorri, H. et al., 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In Proceedings of CHI’17.

Luger, E. & Sellen, A., 2016. Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5286–5297.

Acknowledgements

Thank you to Anja Thieme, Kevin Doherty, Alex Taylor, Saqib Abernut, Peter Bosher and Akito van Troyer for their help and support

Morrison, C. et al., 2017. Imagining Artificial Intelligence Applications with People with Visual Disabilities using Tactile Ideation , pp. 81-90. ACM, 2017. In Proceedings of the 2017 SIGACCESS Conference on Computers and Accessibility. p. 81.

Murray, L. et al., 2016. Capturing social cues with imaging glasses. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 968–972.

Panchanathan, S. Chakraborty, S. & McDaniel, T., 2016. Social Interaction Assistant: A Person-Centered Approach to Enrich Social Interactions for Individuals With Visual Impairments. IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING,, 10(5).

Pradhan, A., Mehta, K. & Findlater, L., 2018. Accessibility Came by Accident: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 459.

Qiu, S., Rauterberg, M. & Hu, J., 2016. Tactile Band: Accessing Gaze Signals from the Sighted in Face-to-Face Communication. In Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. pp. 556–562 .

Sato, D. et al., 2017. Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In Proceedings of the 2017 SIGACCESS Conference on Computers and Accessibility. p. 270

Thieme, A. et al., 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. p. 203.

Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., and Yarin, P. Ambient displays: Turning architectural space into an interface between people and digital information. In International Workshop on Cooperative Buildings, Springer (1998), 22–32.

Wu, S. et al., 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Networking Service. In Proceedings of the Conference on Computer-Supported Cooperative Work and Social Computing.

Ye, Hanlu, et al. "Current and future mobile and wearable device use by people with visual impairments." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2014

Zhao, Y. et al., 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. In Proceedings of the 2018 CHI

Zhao, Y., Bennett, C.L., et al., 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 116.