A Social-Aware Assistant to Support Individuals with Visual Impairments during Social Interaction: A Systematic Requirements Analysis

María Elena Meza-de-Luna[a], Juan R. Terven[b], Bogdan Raducanu[c], Joaquín Salas[d,1]

[a] *Universidad Autónoma de Querétaro, México*
[b] *AiFi Inc., United States*
[c] *Centro de Visión por Computadora, Spain*
[d] *Instituto Politécnico Nacional, México*

Abstract

Visual impairment affects the normal course of activities in everyday life including mobility, education, employment, and social interaction. Most of the existing technical solutions devoted to empowering the visually impaired people are in the areas of navigation (obstacle avoidance), access to printed information and object recognition. Less effort has been dedicated so far in developing solutions to support social interactions. In this paper, we introduce a Social-Aware Assistant (SAA) that provides visually impaired people with cues to enhance their face-to-face conversations. The system consists of a perceptive component (represented by smartglasses with an embedded video camera) and a feedback component (represented by a haptic belt). When the vision system detects a head nodding, the belt vibrates, thus suggesting the user to replicate (mirror) the gesture. In our experiments, sighted persons interacted with blind people wearing the SAA. We instructed the former to mirror the noddings according to the vibratory signal, while the latter interacted naturally. After the face-to-face conversation, the participants had an interview to express their experience regarding the use of this new technological assistant. With the data collected during the experiment, we have assessed quantitatively and qualitatively the

*Send correspondence to: Joaquín Salas. Cerro Blanco 141, Colinas del Cimatario, Querétaro, México, 76090, jsalasr@ipn.mx

device usefulness and user satisfaction.

Keywords: Computer Vision, Visually Impaired, Social Interaction, Assistive Technology

1. Introduction

The visual impairment affects almost every activity of daily living, including mobility, orientation, education, employment, and social interaction. Despite these limitations, blind people reinforce their other perceptual mechanisms: hearing (Kolarik et al., 2016), tact (Voss et al., 2016) and olfact (Araneda et al., 2016), as a form to compensate for their impairment. Some blind people are even capable of developing echolocation, which is the ability to detect objects in their environment by actively creating sounds, *e.g.*, by tapping their canes, lightly stomping their foot, snapping their fingers, or making clicking noises with their mouths (Secchi et al., 2017), and sensing echoes from sound bouncing on the objects. A skilled blind pedestrian can approach an intersection, listen to the traffic, and from audible information alone, judge the spatial layout of intersecting streets, the width of the road, the traffic volume, and the presence of pedestrian islands or medians (Liu and Sun, 2006). Likewise, blind people tend to develop a highly sensitive tactile sense that allows them to read Braille system (Russomanno et al., 2015), and to scan their surroundings (Goldreich and Kanics, 2003). However, even though blind people tend to compensate their lack of sight by augmenting their other senses, the sensory bandwidth of the visual channel is orders of magnitude greater than that of auditory and touch (Loomis et al., 2012).

Nowadays, we witness an unprecedented effort of the scientific community to develop solutions which could restore, even partially, the sense of sight. As a result, several assistive systems using computer vision have been developed for different applications: access to printed information (Cutter and Manduchi, 2013, 2015), object recognition (Gallo and Manduchi, 2011; Guida et al., 2011; Winlock et al., 2010), navigation (Flores et al., 2014; Vera et al., 2013) and social interaction (Krishna and Panchanathan, 2010; Krishna et al., 2010). A comprehensive survey of these applications could be found in Manduchi and Coughlan (2012). Of all these areas, the least exploited one is social interaction (Loomis et al., 2012). The importance of the visual channel is reflected in the type of information exchanged with our counterparts during social interactions. According to Ambady *et al.* (1992), nearly 65% of the generated volume is through non-verbal cues, *i.e.*, eye gaze, facial/hand/body gestures. Thus, a visually impaired person suffers a serious limitation in accessing this rich flow of information and in consequence this deprivation could lead to social isolation, depression, loneliness, and anxiety (Hinds et al., 2003).

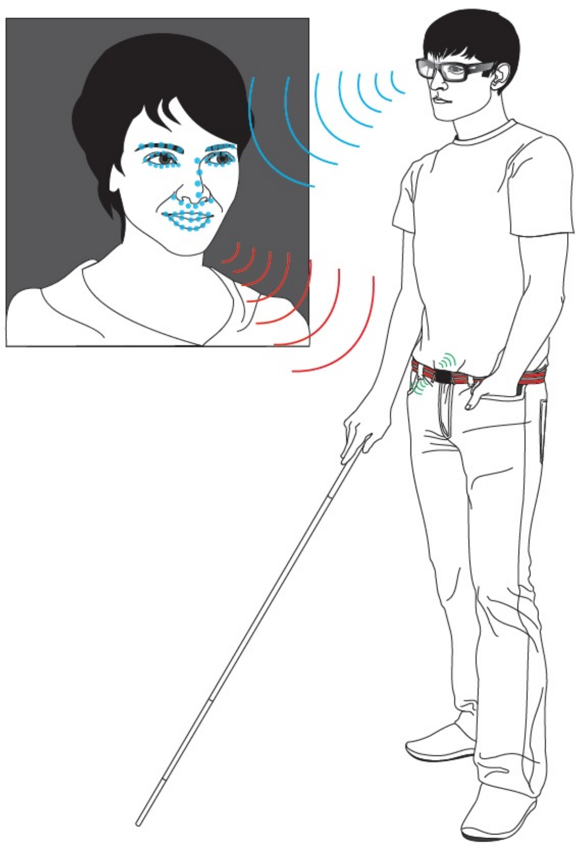
In this work, we develop a wearable technological assistant to provide visually impaired people with cues that may enhance their social interaction during face-to-face conversations. The system architecture is depicted in Figure 1(a). It consists of a perceptive component (represented by smartglasses which have embedded a video camera shown in Figure 1(b)) and a reactive component (represented by a haptic belt shown in Figure 1(c)). The video stream captured by the camera is sent to a computer, where a specialized software runs face detection, tracking, and head gestures algorithms. When a head nodding is detected, this information is sent to the vibratory belt informing the user, who, on his turn, could replicate the gesture. By closing this gesture loop, a mirroring event is triggered.

The main contributions of our approach are:

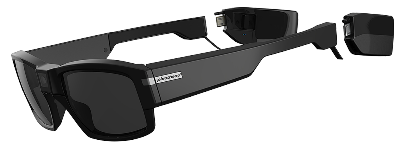
- The introduction a Socially-Aware Assistant (SAA), consisting of perceptive, and feedback components.

- The implementation of a computer-vision-based application for a social assistant on a wearable platform. The use of smartglasses confers the system a first-person perspective, which is ideal for this type of applications when compared with other assistive systems that offer a third-person perspective. R4.03

- The development of new insights obtained by testing the SAA with visually



(a)



(b) Smartglasses

(c) Vibratory Belt

Figure 1: System overview: The smartglasses shown in (b) transmit live video to a computer, which processes the data and sends commands to a vibratory belt shown in (c). The belt receives the commands from the computer and vibrates to provide feedback to the user shown in (a) (Smartglasses with live-stream video capabilities by Pivothead Inc., with permission)

impaired people. We validated our system in a simulated tourism office environment, where we envisioned a face-to-face conversation taking place between a blind person (acting as a *tourist guide*) and a sighted person (acting as *tourists*). As a result, a custom data corpus has been recorded, annotated, and will be made available to the research community upon request.

Our application integrates work we have previously done on head-gesture recognition (Terven et al., 2014, 2016) and social interaction (Meza-de Luna et al., 2016), and new results which include the development of a real-time technology assistant for visual impaired social interaction, and the use of a vibratory belt for haptic feedback.

The rest of the paper is structured as follows: we dedicate section 2 to review the related work in the area of assistive technology for social interaction. In section 3, we describe the system's architecture. In section 4, we explain the design of the experimental study. In section 5, we report our results and discuss the findings of our study. Finally, we draw our conclusion and provide some guidelines for future work.

2. Related Literature

To a large extent, assistive technology for visually impaired people is based on ultrasonic, infrared, or laser sensors (Dakopoulos and Bourbakis, 2010). However, as sighted people acquire most of their information through visual perception, it is tempting to use computer vision to achieve the same goal (Chessa et al., 2016). Nonetheless, limitations in computing power and the lack of reliable algorithms have been a recurrent problem for computer vision in this area. Fortunately, parallel to the revolution in computing technology observed in the last decade, wearable devices are now capable of running real-time vision algorithms (Niu et al., 2017). Comparatively, assistive technologies to support social interactions have been scarcely explored. Our concern is that during a social interaction, a substantial part of the communication comes in the form of non-verbal cues, including face gestures, head movements, gaze direction and others alike (Massé et al., 2017), to which blind people have no access. This represents a severe limitation that may lead to social isolation (Kizar et al., 2016).

The motivation for our work comes from the psychological research devoted to the study of the performance of visually impaired people during social interaction. As Grove *et al.* (1991) observed, when visually impaired people take the initiative in a conversation, their sighted counterparts are less apprehensive, more talkative, and feel more confident since they do not have to worry for a situation that might seem embarrassing. Overall, they feel more satisfied with the interaction (Hecht, 1978), inclined to have more positive thoughts, and even feel empathy with the initiator (McCroskey and McCain, 1974). In another study, Frame (2004) found that, for a successful social interaction, the visually impaired person must be able to adapt to, and perform competently within, his or her assigned role, establish and maintain mutually satisfying relationships with other people, and know how to manage the potential responses.

Hayden (2014) reported a computer-vision based wearable assistant to support the social interaction of blind people. His system consists of a mobile phone, which acts as both a perceptive device through its camera and a computational unit, and a smart watch, which acts as a haptic feedback interface. His wearable assistant recognizes faces and can identify persons passing by and inform the user, in case she/he wants to initiate a conversation. A similar system, intended as a wearable assistant able to characterize different types of social interactions, has been reported in Fathi *et al.* (2012). The assistant employs a head-mounted GoPro camera as a wearable device, and its basic functionality relies on the detection and estimation of people's faces in a group. Based on this information, the role of each person in a group is inferred (either speaker or listener).

Perhaps the most sustained effort so far comes through *iCARE Social Interaction*, a project from Arizona State University. Its goal is to allow blind people to access visual information during social encounters. Through an on-line survey, they detected eight social needs for individuals who are blind and visually impaired (Krishna and Panchanathan, 2010). The most significant one corresponds to feedback on their body mannerism and how it was affecting people's social interactions. The survey concluded that the important needs for the visual impaired include their access to body mannerisms, facial expressions, identity, eye gaze, proxemics and appearance of their social interaction partners, in that order. The prototype they developed uses a camera attached to eyeglasses that communicate with a smartphone. By using computer vision algorithms, the iCARE detects the position of the other person and gives this information to the user through a belt with vibrators (McDaniel et al., 2008). The system can also identify seven basic emotions (happiness, sadness, surprise, anger, fear, disgust, and neutral) and provide this information to the user through a glove made of 14 small vibratory motors (Krishna et al., 2010). The previous results have been combined and presented as an integrated system in Panchanathan *et al.* (2016). Similar to our work, Anam *et al.*(2014) introduced an assistive device based on Google Glasses. Their system recognizes gestures such as Smile, OpenSmile, Sleepy, Yawn, Looking up/down, Looking left/right, and provides speech feedback to the user. By contrast, we chose haptic over auditory feedback, as previous research has identified its advantages (Flores et al., 2015), including its non interference with the user's awareness of other situational events and the reduced cognitive load it requires to recognize messages. Furthermore, our visual system recognizes six gestures, including Looking up/down, Turning left/right, Shaking, and Nodding. However, based on a pre-test inquiry performed with visually impaired people, we concluded that it is easier for them to focus on a single feedback, and opted to use only nodding gestures to obtain insights into the effect of assistive technology for social interaction.

R3.04

The relevance of the present work is that we apply our computer vision algorithms on real video data, captured with the smartglasses, and we test our system in a realistic scenario with visually impaired people.

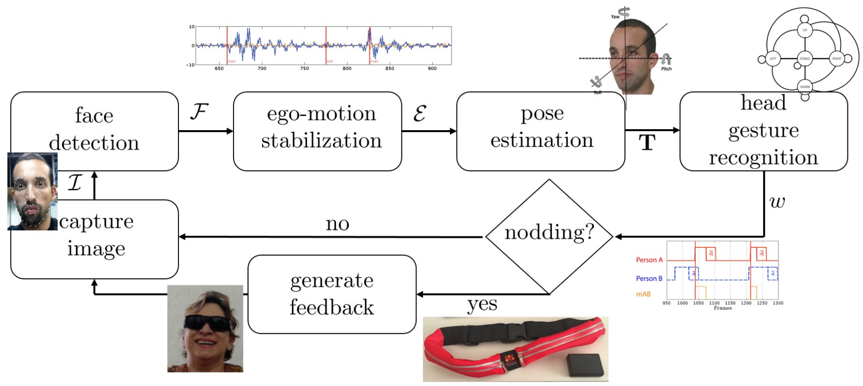


Figure 2: System Architecture: Our system provides haptic feedback for visually impaired users after an interlocutor nodding has been detected. The method to detect nodding involves capturing images $\mathcal{I}$ from the live-stream of smartglasses. From the video, we detect the interlocutor's face. Using the features detected in the scene, $\mathcal{F}$, compensate for ego-motion, $\mathcal{E}$, to estimate the pose, $\mathbf{T}$, and finally recognize the interlocutor's head gesture, w. When appropriate, our system sends commands to a vibratory belt, which in turn provides feedback to the user.

3. System Architecture

Our system's architecture consists of a perceptual component represented by a wearable camera embedded in smartglasses (to detect the head noddings) and a feedback component represented by a haptic belt (to inform the blind user when such an event has been detected) (see Figure 2). The pipeline for our head-nodding detection algorithm consists of the following steps: face detection, face-tracking and stabilization, head pose estimation and head gesture recognition. We explain these elements in the following subsections.

3.1. Face Detection, Tracking and Stabilization

The system activates when the Viola-Jones (Viola and Jones, 2001) detects a face. Once a face is detected, we extract the facial features using non-rigid face tracking with the Supervised Descent Method (SDM) (Xiong and De la Torre, 2013). Our method stabilizes the camera motion (ego-motion) by fitting a motion model to the camera using background features(Terven et al., 2016). R4.04

To estimate camera motion, we extract sparse features of the background and track them using optical flow. We discard features inside the face region to prevent the use of head motion as background motion. More formally, we created a set of n pairs of matched features, $\mathcal{S} = \{(\mathbf{x}_\kappa, \mathbf{x}_{\kappa-1})_j\}$, for $j = 1, \ldots, n$ from the previous and current frames, and where $\mathbf{x}_\kappa^T = (x_\kappa, y_\kappa, 1)$. Using this set of features, we estimated the interframe motion represented as a two-dimensional linear model with four parameters similar to the one proposed in Battiato *et al.* (2007):

$$\mathbf{x}_k = \begin{bmatrix} \lambda\mathbf{R}(\phi) & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x}_{k-1} = \mathbf{T}\mathbf{x}_{k-1}, \tag{1}$$

where $\mathbf{R}$ is a 2×2 rotation matrix, ϕ is the rotation angle, $\mathbf{t}^T = (t_x, t_y)$ is the translation in the x and y direction, $\mathbf{0}$ is a 2×1 vector of zeros, and λ is a scale parameter.

We process the background features removing outliers using localized RANSAC as described by Grundmann *et al.* (2011) to obtain a set of m feature pairs, which we used to solve for the camera motion $\mathbf{T}$ (1) using linear Least Squares. $\mathbf{T}$ establishes a relationship between the facial features with camera motion $\mathbf{x}^w = [x^w, y^w, 1]^T$ and the motionless facial features $\mathbf{x}^s = [x^s, y^s, 1]^T$ as:

$$\mathbf{x}^w = \mathbf{T}\mathbf{x}^s. \tag{2}$$

3.2. Head Pose Estimation

Head pose estimation refers to the computation of a person's head orientation and position (six degrees of freedom) with respect to a camera (Murphy-Chutorian and Trivedi, 2009). In our system, we compute the head pose as follows. For each tracked and stabilized head position, we estimate its pose following the approach described in Martins *et al.* (2008), where 2D image points and 3D model points are matched using the *Pose from Orthography and Scaling with Iterations* method (POSIT) (Dementhon and Davis, 1995). POSIT is

R4.04

a classic algorithm for finding the pose of a 3D model with respect to a camera given matching pairs of 2D image points and 3D object points. For the 3D points, we use a 3D anthropometric model available online in (Martins and Batista, 2008).

3.3. *Head Gestures Recognition*

Once estimated the head pose, the next step is to recognize head gestures such as nodding and shaking. For this purpose, we introduced in Terven et al. (2014) an approach using Hidden Markov Models (HMMs), which works as follows. To train and test the HMMs, we collected a custom dataset using static and wearable cameras. The dataset contains around 150 samples of each of the six gestures. Each gesture in the dataset was translated into a sequence or time series of 20 digits long containing the changes in yaw and pitch in consecutive frames.

R4.04 Typical sequences exhibit larger changes in pitch than in yaw for a nodding gesture and larger changes in yaw than in pitch for a shaking gesture. Inspired by these changes, we define five states in our HMMs and train to recognize six gestures: *nodding*, *shaking*, *looking up*, *looking down*, *turning left*, *turning right*. Although for the present paper we only use *nodding*, the rest of the gestures are included in the model to improve the recognition of similar gestures. R3.03 Our head nodding classifier achieved an Area Under the ROC Curve (AUC) of 0.91, as reported in (Terven et al., 2014, 2016).

3.4. *Haptic Feedback*

Over the past 30 years, there has been a significant amount of research to understand auditory and haptic perception for assisting blind people. One approach has been the development of *general purpose* sensory substitution either by touch (Bach-y Rita, 1967; Collins, 1970) or audition (Capelle et al., 1998; Meijer, 1992). General substitution by touch has led to what is called *distal attribution* or *externalization* (Auvray et al., 2005; Auvray and Myin, 2009; Siegle and Warren, 2010), which refers to sensing tactile stimulation on the skin of objects external to the user. However, general-purpose sensory substitution devices have not proved to be robust and reliable for everyday life and have led to *specific purpose* devices that enable specific tasks such as mobility and orientation.

Even though blind people tend to develop augmented senses, the sensory bandwidth of vision is orders of magnitude greater than that of auditory and touch (Loomis et al., 2012). Because of this, it is necessary to pre-process the visual information to extract only relevant cues and provide high-level feedback.

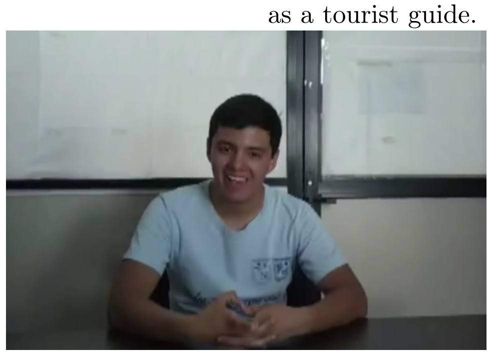
Feedback is an essential component of every assistive technology device. In the case of technology for blind users, the feedback can be acoustic or haptic. Acoustic feedback can be used to provide information about events, the presence of people or animals and to estimate distances (Hill et al., 1993). Haptic feedback refers to touch and tactile sense. In the case of visually impaired people, the widest use of haptic feedback is the white cane which is used to scan the immediate surroundings (Nichols, 1995). Another typical example of haptic feedback is the use of the palm of the hands and fingers to recognize shape and texture of objects. Blind people also use the feet soles to gather information about the surface. In both cases, user acceptance is a key issue.

Our choice for haptic feedback is motivated by three factors: (1) to be non-distracting, (2) to guarantee privacy, and (3) to have good aesthetics. That is, we meant to provide the necessary information without distracting the interlocutor, nor blocking the user's hearing, in a private manner to the user, comfortable and visually appealing. For this, we use a comfortable sports belt (shown in Figure 1(c)) augmented with a custom electronic circuit located at the center of the belt, and two small vibratory motors located on each side of the belt. The electronic circuit contains a digital signal controller to handle the motors and a Bluetooth transceiver to communicate with the computer.

When a head gesture is detected, the computer sends a Bluetooth command to the belt, which in turn decodes the command and activates one of two small vibratory motors to provide feedback to the user. The vibration duration and intensity can be customized to adapt to the user's preferences.

(a) Normal sighted person acting as a tourist.

(b) Visually impaired person acting as a tourist guide.



(c)

Figure 3: Examples of images from our experimental setup. Subfigures (a) and (b) show images obtained from the static cameras placed on the table, and subfigure (c) shows an image taken from the wearable camera used by the visually impaired person. Image resolution is 640 × 480 pixels.

The system was implemented in C++ and takes around 60ms to run all steps achieving a frame rate of 15 frames/sec on HD video. This performance was measured on a tablet equipped with 8GB of RAM and an Intel Core i5 microprocessor running at 1.9GHz without dedicated GPU.

4. Experimental Design

To assess the usefulness of the SAA, we designed an experiment integrating mixed methods (Hernández et al., 2010). To facilitate the conversation between two unknown people, we set up a scenario of a tourist office where the blind person could explain freely what can be visited in the city. The scene offered an accessible and motivating conversation topic for people, regardless of their level of education. To achieve this, we selected participants among the ones who knew the city of Querétaro before they were diagnosed with blindness and could make at least three recommendations on what to visit. In our experiment, blind people (acting as *tourist guides*) and sighted students (acting as *tourists* asking for advice while visiting a city) interacted during three minutes (see Figure 3). Each *tourist* had two conversations with a different *tourist guide*. One of them wearing the SAA (condition $\mathcal{C}_1$) and the other without it (condition $\mathcal{C}_0$). The feedback component generated two different vibrations when the interlocutor's head movement was detected: Vibration on the left side when detecting shaking and vibration on the right side when detecting nodding. It is important to stress that the analysis in this paper is based exclusively on the occurrence of noddings. Therefore, we instructed the *tourist guides* in condition $\mathcal{C}_1$ to mirror the head gestures according to the vibratory signal that was being received at the belt. *Tourist guides* in condition $\mathcal{C}_0$ used similar smartglasses but without receiving feedback. After the experiment, the participants assessed their satisfaction on the interaction using a written questionnaire and a semi-structured interview. We evaluated features such as *interest* (an attitude to be orientated towards what the other person is doing or saying (Adler et al., 1964)), *warmness* (ability of the other person to make one feel confident (Fiske et al., 2007)), *closeness* [R4.05] (quality of emotional proximity where the person is sensitive to the needs and desires of the interlocutor (Ward and Broniarczyk, 2011)), *friendliness* as [R4.08] the ability to show camaraderie; goodwill and light-hearted rapport (American Psychological Association, 2018) and *satisfaction* as the subjective evaluation, of pleasure or disappointment, on the conversation.

4.1. Participants

Our inclusion criteria to participate in the experiment were being at least 18 years old and to signing a participation agreement. All *tourists* were sighted undergraduate students. As our motivation to develop the SAA was to help [R3.01] to curve the period of adaptation, we set as an inclusion criterion for *tourist guides* of having at most three years of having been declared legally blind (visual acuity of 20/200 or less), and totally blind (no light perception) at the time of participation in the experiment. This threshold agrees with previous studies which establish the period of psychological adaptation of blind people Fitzgerald et al. (1987); Robertson et al. (2006) and with the empirical observations carried out by teachers and psychologists caring for the blind participants. R3.02

For the experiment, we made a pre-selection of participants matching people with similar age, verbal and gestural expression. We assessed it during an interview, with two social sciences specialists, by scoring the number of gestures of the blind person, their tendency to direct the face towards the interlocutor, and whether their verbal expressiveness was detailed or succinct. The selected candidates were divided in pairs of *tourists* interacting with the same set of *tourist guides*. The participants in the experiment consisted of 32 volunteer students (50% women), ranging from 18 to 31 years ($\mu = 21.1$, $\sigma = 3.3$). Additionally, there were eight blind people (50% women), ranging from 29 to 67 years ($\mu = 48.6$, $\sigma = 11.6$). The *tourist guides* were invited to participate through a local school specialized in blind people, to evaluate our SAA.

4.2. Development of the Experiment

During the experiment, the *tourists* were not informed of the fact that the *tourist guides* were wearing or not the SAA, nor were they aware of the SAA's functionality. To decrease the prime exposure bias, half of the *tourists* were presented first to a *tourist guide* in condition $\mathcal{C}_1$ and then to a *tourist guide* in condition $\mathcal{C}_0$. For the other half, the protocol was the other way around. Before the experiment, participants completed their general data and an informed consent form.

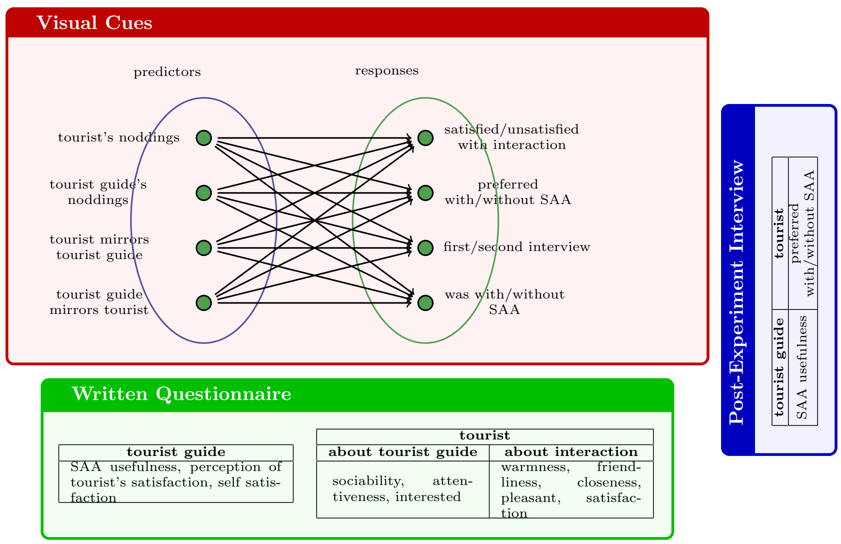
In the written questionnaire, we asked the *tourists* to evaluate the conversation on a scale between 0 and 10 (where 10 is best), inclusive, with Semantic Differential rating based on attributes of the Two-Dimensional Social Interaction Scale (Wai and Bond, 2001) (coldness/warmness, malevolence/friendliness, distant/close, unpleasant/pleasant, dissatisfaction/satisfaction) and their perceptions of the *tourist guides* (retracted/sociability, apathetic/attentiveness, in-different/interested). In the semi-structured interview, the *tourists* were asked to describe their impressions throughout the conversation and to provide a general assessment of their satisfaction. After the second conversation, they were invited to compare both interactions.

The *tourist guides* had two sessions of training using the technology assistant. In the first training session, the blind persons became familiar with the device. This session allowed us to evaluate their ability to interpret feedback. Afterwards, we interviewed them to assess their interacting style and their potential to recommending at least three places to visit. In the second training session, we reinforced the vibration-gesture association to help the blind people to get confidence in their role as tourist guides. In general, each training session lasted about half an hour.

For *tourist guides* in condition $\mathcal{C}_1$, in the experiment, we asked them to nod in response to the haptic stimuli. After each session, the *tourist guide* had a semi-structured interview. During the interview, they described their impressions throughout the conversation with the *tourist*. They evaluated on a scale from 0 (very unsatisfied) to 10 (very satisfied) their satisfaction and the satisfaction they believe that the *tourist* may had in the conversation. For our purposes, we defined *satisfaction in the conversation* as the *subjective evaluation, of pleasure or disappointment, when comparing the self-perception of performance relative to the self-expectations.* Overall, each *tourist guide* participated in four sessions in condition $\mathcal{C}_1$ and four sessions in condition $\mathcal{C}_0$. At the end of the sessions, we asked *tourist guides* to assess the possible difference in the conversations with/without the SAA, its potential usefulness in daily life and possible improvements in future versions.

This study followed ethical standards as stipulated by the American Psychological Association (Flavio et al., 2010). Participants signed an informed consent letter. Confidentiality and person's anonymity were maintained at all times. All video and audio recordings were made with participant's written authorization. The Ethics Committee of the Universidad Autónoma de Querétaro approved the protocol. The experiment took place between February and May

Table 1: Analysis of the Results. In our analysis of the results we determine to what degree the responses can be inferred from the measurements obtained from the vision system through a series of classifiers. A written questionnaire asks the tourists and tourist guides to evaluate different aspects of the conversation. We use a post-experiment interview to find out the perceived usefulness of the SAA. This qualitative evaluation is used as ground truth to both evaluate the performance of a set of classifiers constructed to evaluate the vision system and for the importance analysis of the features in the written questionnaire.



2016.

As a result of the experiment, we created a corpus of data, which for each conversation in $\mathcal{C}_1$ contains three videos (two from the static cameras focused on the interlocutors and one from the *tourist guides*' wearable camera), and for each conversation in $\mathcal{C}_0$ contains two videos (from the static cameras), a written questionnaire answered by the *tourist*, and two recorded qualitative interviews (one with the *tourist guide* and one with the *tourist*).

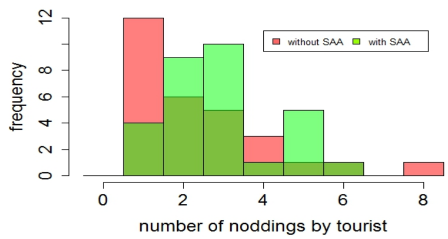
5. Experimental Results

In the experiment described in section §4, visually impaired people acted as *tourist guides* while normally sighted people acted as *tourists* (See Figure 3). In this section, we analyze the result of the conversations. As our objective is to find out the forms the SAA improves the communication with blind people, we search for the strength of the relationship between the outcome of the inferences made with the visual cues we obtain during the conversation, and the responses of the participants to written questionnaires and post-experiment interviews (see Table 1). R4.06

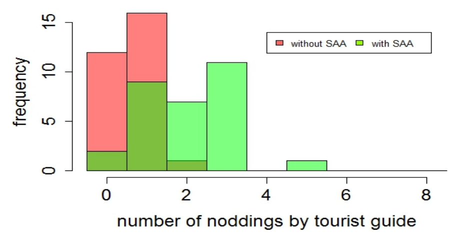
5.1. Visual Cues

In this section, the purpose is to assess whether the number of head noddings or mirroring events generated by either the *tourist* or the *tourist guide* using the SAA (condition $\mathcal{C}_1$) or not using the SAA (condition $\mathcal{C}_0$) shade light on the conversation. Our approach consisted of constructing classifiers to distinguish the responses out of the predictors. The visual cues we used as predictors include the number of noddings made by the *tourist* and the *tourist guide*, and the number of times either the *tourist* mirrored the *tourist guide* or vice-versa. The responses of our classifiers included the satisfaction in the conversation, whether the interview was the first one or the second one the *tourist guide* wearing the SAA or not, and whether the interlocutors (in their words) preferred the conversation with or without the SAA. Using these as the segregation criteria, we divided our data using the predictors. See, for instance, Figure 4(a)-(d), R4.07 which illustrates the case where the response is whether the *tourist guide* was using the SAA or not for different predictors.

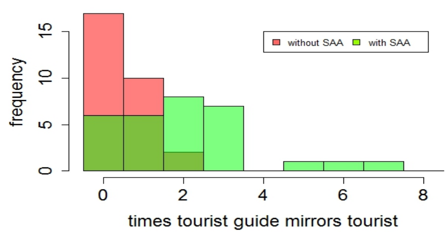
To assess the importance of each predictor in the classification tasks, we performed a feature selection analysis using the Boruta algorithm (Kursa and Rudnicki, 2010), a wrapper based on the Random Forests classifier (Breiman, 2001). In the Boruta algorithm, a feature is considered important whenever its removal degrades the performance of the classifier. To operate, the Boruta algorithm makes copies of the predictors into so-called *shadow variables*, arranging their values in random order. Then, a random forest model is fitted to the extended set of predictors. The importance of each predictor is related to the classifier loss of accuracy when it is removed. The Z-score is the average loss



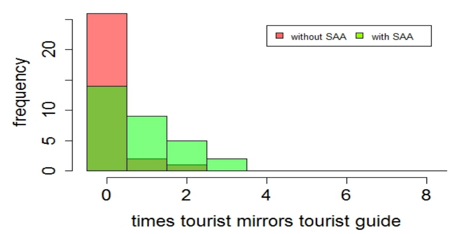
(a)



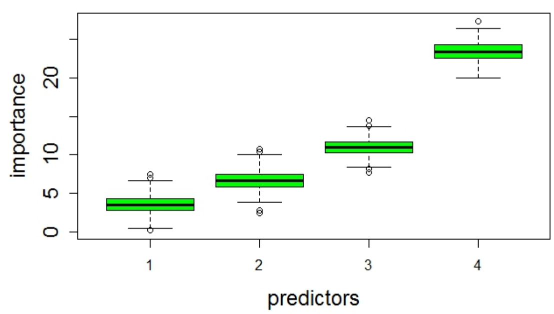
(b)



(c)



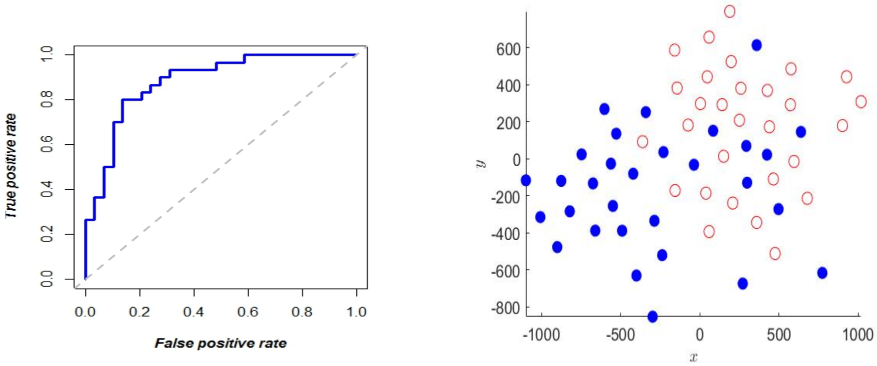
(d)



1. **noddings by tourist**
2. *tourist* **mirrors** *tourist guide*
3. *tourist guide* **mirrors** *tourist*
4. **noddings by** *tourist guide*

(e)

Figure 4: Detection of experimental condition where the visual impaired was using (green) and not using(red) the SAA. Histograms of the number of measured head noddings (a)-(b), and mirroring events (c)-(d), for the detection of the experimental condition where the tourist guide is wearing the SAA (green) and is not wearing the SAA (red) (overlap is shown as reddish green). The output of the Boruta algorithm (e) shows that the measured visual cues are found important with corresponding values 3.42, 6.7, 10.9 and 23.6.



(a) AUC = 0.88 (b) t-SNE

Figure 5: The performance of the classifier to detect the operating condition with the SAA or without it results in the ROC curve in (a). The predictors include the number of noddings and the mirroring events. Together with the response, this spans a five-dimensional classification space. A t-SNE (Maaten and Hinton, 2008) visualization is shown in (b).

of accuracy over all trees in the forest divided by the standard deviation. The variables that are found important are those in which the Z-score is above the maximum Z-score of the shadow variables.

Based on the output of the Boruta algorithm, the visual cues were found useful to distinguish whether the *tourist guide* was using the SAA. The relative importance assigned by the Boruta algorithm was 23.6 for the number of noddings made by the *tourist guide*, 10.9 for the number of times the *tourist guide* mirrored the *tourist*, 6.7 for the number of times the *tourist* mirrored the *tourist guide* and 3.42 for the number of noddings made by the *tourists*. Nonetheless, the visual cues were found not significant to assess: whether the interlocutors were satisfied as a result of the conversation, whether the *tourist* interacted first with a *tourist guide* wearing the SAA or vice-versa, or whether the participants explicitly preferred the conversation with or without the SAA.

Using the Random Forest classifier output by the Boruta algorithm, the visual cues as predictors and whether the *tourist guides* were using the SAA or not, as response, we performed a Receiving Operating Characteristic (ROC) analysis (Swets et al., 2000) Varying the decision threshold τ over the range of the distributions, we computed the number of true positives (*tp*), false positives (*fp*), true negatives (*tn*) and false negatives (*fn*). Then, we evaluated the sensitivity (*tpr*) and specificity (1 - *fpr*) for the ROC curve using

$$tpr = \frac{tp}{tp + fn}, \text{ and } fpr = \frac{fp}{fp + tn}. \tag{3}$$

The area under the ROC curve, illustrated in Figure 5)(a), is 0.88, which suggests that both classes can be distinguished with high confidence. To gain some intuition on the difficulty of the classification problem, we perform an exercise of dimensionality reduction using t-SNE (Maaten and Hinton, 2008) (see Figure 5(b)). The resulting distribution of features in a bidimensional space coincides intuitively with the outcome of our previous analysis.

5.2. Written Questionnaire

The written questionnaire evaluated the personal impressions of the interlocutors regarding the conversation.

When invited to evaluate quantitatively their experience, the *tourist guides* scored the usefulness of the SAA, their perception of the *tourist* satisfaction, and their satisfaction with a median value of 10, 9 and 9, respectively. At the end, we asked the *tourists* to evaluate the conversation in terms of *warmness*, *friendliness*, *closeness*, *pleasant*, and *satisfaction*. They also evaluated their perception of the *tourist guide* in terms of *sociability*, *attentiveness*, and *interest* in the conversation. In general, the distributions of the evaluation for these concepts are high or left-skewed.

To evaluate whether there are statistically significant differences in the conversation between the two conditions (with or without the SAA), we first notice that the distributions are severely skewed toward high values. In these cases, a common practice is to use the Wilcoxon Signed Rank test (Wilcoxon, 1945) as an alternative to the dependent samples t-test. In our decision rule, we used a significance value of $\alpha = 0.05$, *i.e.*, if $p \leq 0.05$, the groups differ significantly, and the device makes a difference in the conversation viewed from the *tourist*'s perspective. Table 2 shows the results of applying this test to each element of the conversation scored by the participants. These results show that there are no differences in the satisfaction of the conversation, as perceived by the *tourists*, for the cases when the *tourist guide* was wearing the SAA or not. One explanation for this finding could be the fact that most of the sighted people are not familiar interacting with blind persons and do not know what to expect from such a conversation. These results may not be surprising due to the positive evaluation from most of the participants, leading to accept the null hypothesis that there are no differences between the means of the two groups.

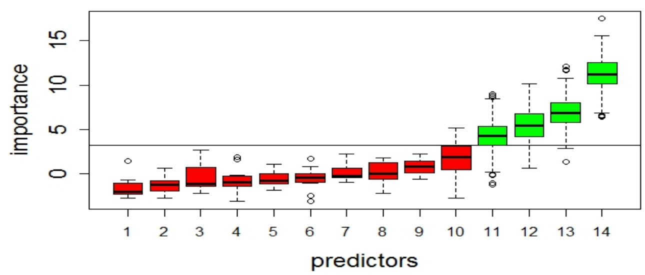
Table 2: Wilcoxon Signed Rank Test results. It is shown that there are no differences in the satisfaction of the conversation, as perceived by the *tourists*, for the cases where the *tourist guide* was wearing the SAA or not.

	Warmness	Friendliness	Closeness	Pleasant	Satisfaction	Sociability	Attentiveness	Interest
V	244	130	200	72	97	239	141	213
p	0.56	0.81	0.49	0.55	0.94	0.41	0.93	0.33

To infer whether there is a subset of predictors which best distinguishes between a good evaluation of the conversation and a bad one, we performed an analysis of feature importance using the Boruta algorithm. In Figure 6, we summarize the results. There the red and green boxplots represent *importance scores* of rejected and confirmed attributes, respectively.

The features found relevant are *interest*, *warmness*, *closeness* and *tourist's* satisfaction with a mean importance of 11.32, 6.94, 5.19 and 4.37 respectively. This value corresponds to the Z-score, which for a given variable is the difference between the mean decrease in accuracy divided by the standard deviation of the loss. The Boruta algorithm estimated an accuracy of 0.68 in assessing satisfaction in the conversation using these four factors alone.

Our results rely on the use of the Boruta algorithm. It may be interesting to explore other feature selection and classification techniques, *e.g., deep learning,* to corroborate our inferences.



1. *Tourist Guide*'s **Perception of** *Tourist* **Satisfaction**
2. **Conversation**
3. *Tourist Guide*'s **Satisfaction**
4. **Condition**
5. **Pleasant**
6. **Satisfaction**
7. **Friendliness**
8. **Attentiveness**
9. **Sociability**
10. *Tourist*'s **Perception of** *Tourist Guide*'s **Satisfaction**
11. *Tourist*'s **Satisfaction**
12. **Closeness**
13. **Warmness**
14. **Interest**

Figure 6: Relevant Feature Selection. Out of the written questionnaire, we selected the answers more relevant to predict the satisfaction in the conversation using the Boruta algorithm. In this illustration, relevant (Tourist's Satisfaction, Closeness, Warmness and Interest) features are above the horizontal line.

5.3. Post-Experiment Interview

Once the conversation between the *tourist* and *tourist guide* ended, we interviewed the *tourist guides* to collect their general appreciation of the SAA's usefulness. This interview focused on identifying positive and negative aspects of the SAA, with the purpose of enhancing and improving their functionality in future versions.

On the upside, seven out of eight *tourist guides* found the SAA useful to reinforce their auditive information, which increases their confidence and comfort during the conversation by confirming that the interlocutor was paying attention to the conversation. For instance, one blind participant said: *"I liked it* R4.09 *because the feedback it provides, reaffirms what I've heard in the conversation."* R4.05 Overall, they seemed excited about using the SAA in their daily life for longer periods of time (days, weeks). The possibility to use it outdoors has been well received.

On the downside, a single tourist guide commented us that the vibratory R4.01 belt was a molesting factor, as he perceived its vibrations as disturbing. In future research, we will explore the mechanisms under which a blind user could recognize the haptic feedback more naturally, perhaps with focused training as Buimer *et al.* (2018) demonstrates. Another aspect worth mentioning is that to run correctly, the SAA requires the *tourist guide* to face the *tourist*. If this is not the case, the SAA will not have the visual information it requires. In our experiments, before the interaction started, we explicitly instructed the *tourist guides* to adjust their pose to this situation. R4.01a R4.05

Despite these comments, the *tourist guides* identified application areas they would like to see included in future versions of the SAA. The list of recommendations includes: the recognition of facial expressions (not only facial emotions but also communicative gestures, such as thinking, smiling, interested, concerned); a description of the interlocutor's appearance (what he/she wears and how she/he is dressed); information about the distance to the interlocutor; an early warning signal activated when a person is nearby. Additionally, the interviewees also mentioned other desired functionalities. Although these functionalities are not related to social interaction, they are mentioned here for the sake of completeness. These include: a description of the scene for travel assistance (*i.e.*, when they have reached a corner, a crossroads, a street sign, an information panel); support in finding objects; support to locate and identify products in the supermarket, etc.

Additionally, some *tourist guides* expressed they felt the need to have more control on the SAA. The current version was designed as a prototype and included mainly functional aspects. New releases of the SAA should consider user studies to improve its design.

5.4. Discussion

It has long been observed that head gestures, such as nodding, increase the opportunities for a person to be liked (McGovern et al., 1979; Gifford et al., 1985), while the occurrence of mirroring is an early predictor of acceptance (Van Baaren et al., 2003; Guéguen et al., 2009; Jacob et al., 2011; Farley, 2014). Our results show that it is possible to predict whether the visually impaired person is using the SAA, based on the measurements obtained by the computer vision system, and the satisfaction in the conversation, based on the written questionnaire. Indeed, the level of performance achieved by our predictors is significative in the context of the social sciences studies (Jayles et al., 2017).

In general, the written evaluation of the participants tends to be high. It seems that for *tourist guides* the SAA was useful to confirm their auditive perception. Sighted people usually have additional information cues (*e.g.* face and hand gestures, body postures, etc.) that help them to evaluate elements of a conversation which are not available for the blind people. Hence, since the SAA provided extra information about head gestures, some *tourist guides* expressed increased confidence to interact with sighted people. Indeed, they found that the SAA's feedback allowed them to assess better whether the interlocutor was attentive to the conversation, a piece of knowledge that is not trivial for them to obtain. For example, a blind person declared that the SAA could be used as a re-training tool for individuals with visual impairments to move their face towards their interlocutor, as, lacking the practice, one can lose this gesture (Deville et al., 2008).

As some research notices (Turkle, 2015), face to face communication is essential to improve social interaction and is critical for the development of empathy. Hence, by losing the habit of looking for the face of their interlocutor, blind people may be downplaying their social interaction and contributing to increasing their isolation and low self-esteem. In another interview, a blind person declared that individuals with visual impairments could find the SAA useful for their communication with speech-impaired people. This statement highlights the importance of developing assistive technology centered in the users as it is easy to overlook their needs.

In the case of the *tourists*, their high evaluation for the conversation seems to be explained mainly by the novel experience it represents for a sighted person to interact with a visually impaired one. Therefore, some benefits resulting from wearing the SAA may have been eclipsed by this positive excitement.

Some of the participants reported that, at first use, the SAA can be a distractor that affects the natural flow of conversation. This effect stresses out the importance of providing enough time to overcome the initial distraction and to adapt to the new stimuli. In turn, this training will help users to integrate the new information to their communicative abilities in the middle of an intense internal process to which they are trying to adapt (Lun Khoo et al., 2013).

Conclusion

This paper introduces a wearable assistant for blind users to support their social interaction. The SAA developed is capable of automatically recognizing head-noddings and to incite mirroring events using visual information obtained from cameras and providing haptic feedback to the blind user. With the inference of satisfaction derived from the qualitative evaluation of the interviews as ground truth, we showed that our computer vision algorithms are capable of recognizing whether the SAA was being used during conversations and whether the blind's interlocutors were satisfied with the conversation using the quantitative evaluation of predictors such as *warmness*, *closeness*, *interest* and *tourist*'s satisfaction.

In the future, our research will focus on the face-to-face communication between a blind person and a person with whom he/she shares an intimate bond: either a relative or a close friend. In this case, we expect that the sighted person will notice a difference in the communication when the blind person wears the SAA. Another research direction is to test our SAA in a multi-party setting. Nevertheless, in this case, the attention could shift from one person to the other, depending on the conversation dynamics. However, the same procedure described in this paper should be applied in this case, too, since in general only one-to-one conversations are possible at any given moment. Additionally, we are planning to extend our system in a variety of forms which include the recognition of additional human gestures and the incorporation of a better feedback system. For the development of assistive technology for the visually impaired, the major challenge seems to be the former given the reduction in information bandwidth

R4.02

from the visual channel to the haptic one. Our selection of nodding as the only visual gesture aims to help us to achieve insight into the problems of providing feedback to the visually impaired. Recent research may guide the development of future feedback systems exploring the limits of haptic discrimination and the combined use of haptics and audio feedback.

Acknowledgments

This work was partially supported by CONACYT under Grant INFR-2016-01: 268542, by IPN-SIP under Grant No. 20170566 and by UCMexus, for Joaquín Salas; and by MINECO Grant TIN2016-79717-R, Spain, for Bogdan Raducanu; Juan Ramón Terven was partially supported by Instituto Tecnológico Nacional de México and CONACYT.

References

Adler A, Linton J, Vaughan R. Social Interest: A Challenge to Mankind. Capricorn Books New York, 1964.

Ambady N, Rosenthal R. Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. Psychological Bulletin 1992;111(2):256–74.

American Psychological Association . APA Dictionary of Psychology. American Psychological Association, 2018.

Anam I, Alam S, Yeasin M. Expression: A Dyadic Conversation Aid using Google Glass for People Who are Blind or Visually Impaired. In: International Conference on Mobile Computing, Applications and Services,. IEEE; 2014. p. 57–64.

Araneda R, Renier L, Rombaux P, Cuevas I, De Volder A. Cortical Plasticity and Olfactory Function in Early Blindness. Frontiers in Systems Neuroscience 2016;10.

Auvray M, Hanneton S, Lenay C, O'Regan K. There is Something Out There: Distal Attribution in Sensory Substitution, Twenty Years Later. Journal of Integrative Neuroscience 2005;4(04):505–21.

Auvray M, Myin E. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. Cognitive Science 2009;33(6):1036–58.

Battiato S, Gallo G, Puglisi G, Scellato S. SIFT Features Tracking for Video Stabilization. In: IEEE International Conference on Image Analysis and Processing. 2007. p. 825–30.

Breiman L. Random Forests. Machine Learning 2001;45(1):5–32.

Buimer H, Bittner M, Kostelijk T, Van Der Geest T, Nemri A, Van Wezel R, Zhao Y. Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device. PloS one 2018;13(3):e0194737.

Capelle C, Trullemans C, Arno P, Veraart C. A Real-Time Experimental Prototype for Enhancement of Vision Rehabilitation using Auditory Substitution. IEEE Transactions on Biomedical Engineering 1998;45(10):1279–93.

Chessa M, Noceti N, Odone F, Solari F, Sosa-García J, Zini L. An Integrated Artificial Vision Framework for Assisting Visually Impaired Users. Computer Vision and Image Understanding 2016;149:209–28.

Collins C. Tactile Television-Mechanical and Electrical Image Projection. IEEE Transactions on Man-Machine Systems 1970;11(1):65–71.

Cutter M, Manduchi R. Real Time Camera Phone Guidance for Compliant Document Image Acquisition Without Sight. In: IEEE International Conference on Document Analysis and Recognition. 2013. p. 408–12.

Cutter M, Manduchi R. Towards Mobile OCR: How To Take a Good Picture of a Document Without Sight. In: International Symposium on Document Engineering. 2015. p. 75–84.

Dakopoulos D, Bourbakis N. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 2010;40(1):25–35.

Dementhon D, Davis L. Model-Based Object Pose in 25 Lines of Code. International Journal of Computer Vision 1995;15(1-2):123–41.

Deville B, Bologna G, Vinckenbosch M, Pun T. Guiding the Focus of Attention of Blind People with Visual Saliency. In: Workshop on Computer Vision Applications for the Visually Impaired. 2008. .

Farley S. Nonverbal Reactions to an Attractive Stranger: The Role of Mimicry in Communicating Preferred Social Distance. Journal of Nonverbal Behavior. Journal of Nonverbal Behavior 2014;38(2):195–208.

Fathi A, Hodgins J, Rehg J. Social Interactions: A First-Person Perspective. In: IEEE Conference on Computer Vision and Pattern Recognition. 2012. p. 1226–33.

Fiske S, Cuddy A, Glick P. Universal Dimensions of Social Cognition: Warmth and Competence. Trends in Cognitive Sciences 2007;11(2):77–83.

Fitzgerald R, Ebert J, Chambers M. Reactions to Blindness: A Four-year Follow-up Study. Perceptual and motor skills 1987;64(2):363–78.

Flavio L, Carneiro M, Madge J, Young M, Meadows D, Johnson G, Fienup J, Quance D, Schaller N. Publication Manual of the American Psychological Association. In: Reports, Results and Recommendations from Technical Events Series. Number C30-56; 2010. .

Flores G, Kurniawan S, Manduchi R, Martinson E, Morales L, Sisbot EA. Vibrotactile Guidance for Wayfinding of Blind Walkers. IEEE Transactions on Haptics 2015;8(3):306–17.

Flores G, Manduchi R, Zenteno E. Ariadne's Thread: Robust Turn Detection for Path Back-Tracing using the iPhone. In: Ubiquitous Positioning Indoor Navigation and Location Based Service. 2014. p. 133–40.

Frame M. Blind Spots: The Communicative Performance of Visual Impairment in Relationships and Social Interaction. Springfield, 2004.

Gallo O, Manduchi R. Reading 1D Barcodes with Mobile Phones using Deformable Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 2011;33(9):1834–43.

Gifford R, Ng C, Wikinson M. Nonverbal Cues in the Employment Interview: Links between Applicant Qualities and Interviewer Judgements. Applied Psychology 1985;70(4):729–36.

Goldreich D, Kanics I. Tactile Acuity is Enhanced in Blindness. Journal of Neuroscience 2003;23(8):3439–45.

Grove T, Werkman D. Conversations with Ablebodied and Visibly Disabled Strangers: An Adversarial Test of Predicted Outcome Value and Uncertainty Reduction Theories. Human Communication Research 1991;17:507–34.

Grundmann M, Kwatra V, Essa I. Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths. In: IEEE Conference on Computer Vision and Pattern Recognition. 2011. p. 225–32.

Guéguen N, Jacob C, Martin A. Mimicry in Social Interaction: Its Effect on Human Judgement and Behavior. European Journal of Social Sciences 2009;8(2):253–9.

Guida C, Comanducci D, Colombo C. Automatic Bus Line Number Localization and Recognition on Mobile Phones. A Computer Vision Aid for the Visually Impaired. In: Image Analysis and Processing. Springer; 2011. p. 323–32.

Hayden D. Wearable-Assisted Social Interaction as Assistive Technology for the Blind. In: Master Thesis. 2014. .

Hecht M. The Conceptualization and Measurement of Interpersonal Communication Satisfaction. Human Communication Research 1978;4:253–64.

Hernández R, Fernández C, Baptista P. Metodología de la Investigación. Mc Graw Hill 2010;.

Hill E, Rieser J, Hill M, Hill M. How Persons with Visual Impairments Explore Novel Spaces: Strategies of Good and Poor Performers. Journal of Visual Impairment & Blindness 1993;87(8):295–301.

Hinds A, Sinclair A, Park J, Suttie A, Paterson H, Macdonald M. Impact of an Interdisciplinary Low Vision Service on the Quality of Life of Low Vision Patients. British Journal of Ophthalmology 2003;87(11):1391–6.

Jacob C, Guéguen N, Martin A, Boulbry G. Retail Sales people's Mimicry of Customers: Effects on Consumer Behavior. Journal of Retailing and Consumer Services 2011;18(5):381–8.

Jayles B, Kim H, Escobedo R, Cezera S, Blanchet A, Kameda T, Sire C, Theraulaz G. How Social Information can Improve Estimation Accuracy in Human Groups. Proceedings of the National Academy of Sciences 2017;:201703695.

Kizar O, Dalkilic M, Kargun M, Ramazanoglu F, Bayrak M. Comparison of Loneliness Levels in Visually Impaired from Different Sports Branches. The Anthropologist 2016;24(3):853–8.

Kolarik A, Moore B, Zahorik P, Cirstea S, Pardhan S. Auditory Distance Perception in Humans: A Review of Cues, Development, Neuronal Bases, and Effects of Sensory Loss. Attention, Perception, and Psychophysics 2016;78(2):373–95.

Krishna S, Bala S, McDaniel T, McGuire S, Panchanathan S. VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions. In: Extended Abstracts on Human Factors in Computing Systems. ACM; 2010. p. 3637–42.

Krishna S, Panchanathan S. Assistive Technologies as Effective Mediators in Interpersonal Social Interactions for Persons with Visual Disability. In: International Conference on Computers Helping People with Special Needs. Springer; volume 6180 of *LNCS*; 2010. p. 316–23.

Kursa M, Rudnicki W. Feature Selection with the Boruta Package. Journal of Statistical Software 2010;36(11).

Liu J, Sun X. A Survey of Vision Aids for the Blind. In: IEEE Congress on Intelligent Control and Automation. volume 1; 2006. p. 4312–6.

Loomis J, Klatzky R, Giudice N. Sensory Substitution of Vision: Importance of Perceptual and Cognitive Processing. Assistive Technology for Blindness and Low Vision 2012;:161–93.

Lun Khoo W, Knapp J, Palmer F, Ro T, Zhu Z. Designing and Testing Wearable Range-Vibrotactile Devices. Journal of Assistive Technologies 2013;7(2):102–17.

Meza-de Luna ME, Terven J, Raducanu B, Salas J. Assessing the Influence of Mirroring on the Perception of Professional Competence using Wearable Technology. IEEE Transactions on Affective Computing 2016;.

Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research 2008;9:2579–605.

Manduchi R, Coughlan J. (Computer) Vision without Sight. Communications of the ACM 2012;55(1):96–104.

Martins P, Batista J. Monocular Head Pose Estimation. In: International Conference on Image Analysis and Recognition. Springer; volume 5112 of *LNCS*; 2008. p. 357–68. `http://aifi.isr.uc.pt/index.html`.

Massé B, Ba S, Horaud R. Tracking Gaze and Visual Focus of Attention of People Involved in Social Interaction. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017;.

McCroskey J, McCain T. The Measurement of Interpersonal Attraction. Speech Monographs 1974;41:261–6.

McDaniel T, Krishna S, Balasubramanian V, Colbry D, Panchanathan S. Using a Haptic Belt to Convey Non-Verbal Communication Cues During Social

Interactions to Individuals who are Blind. In: Workshop on Haptic Audio visual Environments and Games. IEEE; 2008. p. 13–8.

McGovern T, Jones B, Morris S. Comparison of Professional versus Student Ratings of Job Interviewee Behavior. Journal of Counseling Psychology 1979;26(2):176–9.

Meijer P. An Experimental System for Auditory Image Representations. IEEE Transactions on Biomedical Engineering 1992;39(2):112–21.

Murphy-Chutorian E, Trivedi M. Head Pose Estimation in Computer Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 2009;31(4):607–26.

Nichols A. Why Use The Long White Cane? `http://tinyurl.com/hwbrody`; 1995. Accessed: 2016-04-01.

Niu L, Qian C, Rizzo J, Hudson T, Li Z, Enright S, Sperling E, Conti K, Wong E, Fang Y. A Wearable Assistive Technology for the Visually Impaired with Door Knob Detection and Real-Time Feedback for Hand-to-Handle Manipulation. In: IEEE Computer Vision and Pattern Recognition. 2017. p. 1500–8.

Panchanathan S, Chakraborty S, McDaniel T. Social Interaction Assistant: A Person-Centered Approach to Enrich Social Interactions for Individuals With Visual Impairments. IEEE Journal of Selected Topics in Signal Processing 2016;10(5):942–51.

Bach-y Rita P. Sensory Plasticity. Acta Neurologica Scandinavica 1967;43(4):417–26.

Robertson N, Burden M, Burden A. Psychological morbidity and problems of daily living in people with visual loss and diabetes: do they differ from people without diabetes? Diabetic medicine 2006;23(10):1110–6.

Russomanno A, O'Modhrain S, Gillespie B, Rodger M. Refreshing Refreshable Braille Displays. IEEE Transactions on Haptics 2015;8(3):287–97.

Secchi S, Lauria A, Cellai G. Acoustic Wayfinding: A Method to Measure the Acoustic Contrast of Different Paving Materials for Blind People. Applied Ergonomics 2017;58:435–45.

Siegle J, Warren W. Distal Attribution and Distance Perception in Sensory Substitution. Perception 2010;39(2):208–23.

Swets J, Dawes R, Monahan J. Better Decisions through Science. Scientific American 2000;283:82–7.

Terven J, Raducanu B, Meza-de Luna ME, Salas J. Head-Gestures Mirroring Detection in Dyadic Social Interactions with Computer Vision-Based Wearable Devices. Neurocomputing 2016;175:866–76.

Terven J, Salas J, Raducanu B. Robust Head Gestures Recognition for Assistive Technology. In: LNCS. Springer; volume 8495; 2014. p. 152–61.

Turkle S. Reclaiming Conversation: The Power of Talk in a Digital Age. Penguin, 2015.

Van Baaren R, Holland R, Steenaert B, Knippenberg V. Mimicry for Money: Behavioral Consequences of Imitation. Journal of Experimental Social Psychology 2003;39(4):393–8.

Vera P, Zenteno D, Salas J. A Smartphone-Based Virtual White Cane. Pattern Analysis and Applications 2013;17(3):623–32.

Viola P, Jones M. Robust Real-Time Object Detection. International Journal of Computer Vision 2001;4.

Voss P, Alary F, Lazzouni L, Chapman E, Goldstein R, Bourgoin P, Lepore F. Crossmodal Processing of Haptic Inputs in Sighted and Blind Individuals. Frontiers in Systems Neuroscience 2016;10.

Wai T, Bond A. Development and Validation of the Two-Dimensional Social Interaction Scale. Psychiatry research 2001;103(2):249–60.

Ward M, Broniarczyk S. It's Not Me, It's You: How Gift Giving Creates Giver Identity Threat as a Function of Social Closeness. Journal of Consumer Research 2011;38(1):164–81.

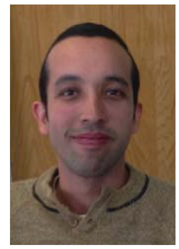
Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1945;1(6):80–3.

Winlock T, Christiansen E, Belongie S. Toward Teal-Time Grocery Detection for the Visually Impaired. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2010. p. 49–56.

Xiong X, De la Torre F. Supervised Descent Method and Its Applications to Face Alignment. In: IEEE Conference on Computer Vision and Pattern Recognition. 2013. p. 532–9. http://www.humansensing.cs.cmu.edu/intraface/.

María-Elena Meza-de-Luna is a researcher at the Universidad Autónoma de Querétaro, México. She is interested in the social and cultural issues to prevent the expression of violence. Currently, she is director of !Atrévete Ya!/iHollaback!-Querétaro, an organization to prevent street harassment (www.atrevete-ya.org) and president of IIPSIS, an NGO devoted to research and intervention in psychosocial matters. Contact her at mezamariel@gmail.com.

Juan R. Terven is a senior research scientist at AiFi Inc. His research interests include computer vision, machine learning, and assistive technologies design. He received a Ph.D. in computer vision from IPN, Mexico. Contact him at `j.r.terven@ieee.org`.



Bogdan Raducanu is a senior researcher and project director at the Computer Vision Center in Barcelona, Spain. His research interests include computer vision, pattern recognition, and social computing. Raducanu received a Ph.D. in computer science from the University of the Basque Country, Bilbao, Spain. Contact him at `bogdan@cvc.uab.es`.

Joaquín Salas is a professor in the area of Computer Vision at Instituto Politécnico Nacional. His professional activity include the development of applications based on computer vision and pattern recognition. Salas received a Ph.D. in computer science from ITESM, México. Contact him at `salas@ieee.org`.