TERESA: A Socially Intelligent Semi-autonomous Telepresence System*

Kyriacos Shiarlis[1], João Messias[1], Maarten van Someren[1], Shimon Whiteson[1],

Jaebok Kim[2], Jered Vroon[2], Gwenn Englebienne[2], Khiet Truong[2], Vanessa Evers[2],

Noé Pérez-Higueras[3], Ignacio Pérez-Hurtado[3], Rafael Ramon-Vigo[3], Fernando Caballero[3], Luis Merino[3],

Jie Shen[4], Stavros Petridis[4], Maja Pantic[4], Lasse Hedman[5], Mårten Scherlund[5], Raphaël Koster[6], Hervé Michel[6]

[1]University of Amsterdam
Amsterdam, Netherlands

[2]University of Twente
Enschede, Netherlands

[3]Pablo de Olavide University
Seville, Spain

[4]Imperial College London
London, United Kingdom

[5]Giraff Technologies AB
Västerås, Sweden

[6]MADoPA, Centre Expert
Rosières-près-Troyes, France

Abstract— **TERESA is a socially intelligent semi-autonomous telepresence system that is currently being developed as part of an FP7-STREP project funded by the European Union. The ultimate goal of the project is to deploy this system in an elderly day centre to allow elderly people to participate in social events even when they are unable to travel to the centre.**

In this paper, we present an overview of our progress on TERESA. We discuss the most significant scientific and technical challenges including: understanding and automatically recognizing social behaviour; defining social norms for the interaction between a telepresence robot and its users; navigating the environment while taking into account social features and constraints; and learning to estimate the social impact of the robot's actions from multiple sources of feedback.

We report on our current progress on each of these challenges, as well as our plans for future work.

Fig. 1: *Left, middle: A user of the TERESA system can remotely interact with others in an elderly day centre. Right: The TERESA telepresence robot. Its most relevant sensors are labeled:*
1 – Dalsa Genie^{TM} HM1400; 2 – VoiceTracker^{TM} II; 3 – Microsoft Kinect^{TM}; 4 – 2x Hokuyo^{TM} UST-10LX; 5 – XSens^{TM} MTi-30 AHRS.

I. INTRODUCTION

In recent years, advances in mechatronics, artificial intelligence, machine learning, computer vision, and signal processing have enabled the application of robotic systems to many real-world environments. Since many of these environments involve humans, developing robotics systems with social intelligence is an increasingly important goal.

In this paper, we introduce the TERESA[1] project, which aims to develop a telepresence robot of unprecedented social intelligence, thereby helping to pave the way for the deployment of robots in settings such as homes, schools, and hospitals that require substantial human interaction. In telepresence systems, a human controller remotely interacts with people by guiding a remotely located robot, allowing the controller to be more physically present than with standard teleconferencing. The TERESA project is developing a new telepresence system that frees the controller from low-level decisions regarding navigation and body pose in social settings. Instead, TERESA will have the social intelligence to perform these functions automatically. The project's main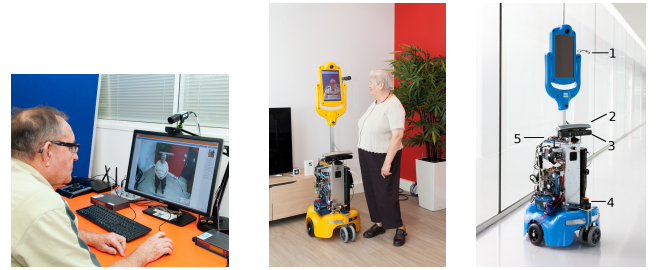 result will be a new partially autonomous telepresence system with the capacity to make socially intelligent low-level decisions for the controller. TERESA will semi-autonomously navigate among groups, maintain face-to-face contact during conversations, and display appropriate body poses. An important feature of the project is that, since it is difficult to manually and explicitly formulate socially acceptable behaviours, these will instead be learned from real data collected under various social situations.

Achieving these goals requires advancing the state of the art in cognitive robotic systems. The project will not only generate new insights into socially normative robot behaviour (see Section II), it will produce new algorithms for detecting and interpreting social behaviour (see Section III), navigating in human-inhabited environments (see Section IV), and modeling social costs using real data from instances of human-telepresence robot interactions (see Section V).

The project culminates in the deployment of TERESA in an elderly day centre. Because such day centres are a primary social outlet, many people become isolated when they cannot travel to them, e.g., due to illness. TERESA's socially intelligent telepresence capabilities will enable them to continue social participation remotely.

A. TERESA's Hardware Features

TERESA is built upon the commercially available Giraff telepresence device, which consists of a video screen at-

tached to a wheeled base. The device can be controlled via a computer by a user in a remote location, whose face appears on the devices video screen (see Figure 1(c)). It has a camera, microphone and a speaker to enable video conferencing.

The TERESA robot has been augmented with additional state-of-the-art hardware. This includes two Hokuyo™ UST 10X laser rangefinders, one XSens™ MTi-30 AHRS gyroscope, one Microsoft Kinect™ sensor, one VoiceTracker™ II array microphone, and a Dalsa Genie™ HM1400/XDR camera. Two extra computers have also been added to process and store all the extra data collected by these sensors. Finally, an extra battery has been added to power the additional equipment.

B. TERESA's Socially-Intelligent Features

1) Social Navigation: A crucial task for TERESA is to be able to navigate its environment in a socially intelligent way. In a typical scenario, the telepresent *visitor* chooses a location of interest from a list situated in a graphical interface and TERESA must navigate from its current location to the assigned location in a non-intrusive and social acceptable manner. For example, if people are crossing the room at that moment, the robot must decide wether to give way or not. It must also decide on an appropriate velocity, so that its behaviour does not appear dangerous.

2) Social Conversation: Another crucial task is to allow seamless and comfortable interaction between the telepresent visitor and *interaction targets*, i.e., the person or group with whom they wish to interact. When the visitor chooses to interact with a certain person, TERESA maintains appropriate eye contact and body pose behaviour to keep the conversation fluent. For example, if the interaction target sits down, TERESA should approach and lower its head so that eye contact can continue during the conversation. If the conversation is within a group, TERESA should continually turn to face the current speaker.

II. SOCIALLY NORMATIVE HUMAN-TELEPRESENCE ROBOT INTERACTION

Improving TERESA's perceived social intelligence requires developing a new, deeper understanding of socially normative interaction between telepresence robots and humans. This involves determining which socially normative telepresence behaviours are most effective in supporting remote social interaction between visitors and interaction targets that is in correspondence with the context, embodiment, and behaviours of the telepresence robot.

A. Observing the Elderly Interact with Each Other

We conducted observational studies with an elderly target group, while they engaged in social activities. We observed a total of six different activities in three locations, ranging from an unscheduled cup of coffee shared by six friends to a celebratory activity with over one hundred attendees.

Our results show that people actively provide each other with space to navigate. For example, during a shared meal, caregivers and elderly without walking problems actively cleared away all walking aids from the main paths. In addition, we found that the great majority of interactions took place between sitting elderly. Interactions between standing or moving elderly were uncommon.

Part of the observed population seems to maintain a reasonable distance from their communication partners, in line with what *proxemics* would predict [6]. However, among people with hearing problems (identifiable by their hearing aids and utterances), a 'leaning' behaviour was commonly observed: during conversation, the person with hearing problems would turn her upper body and lean towards her conversation partner. The partner commonly reciprocated the leaning behaviour. In most cases, the leaning behaviour was only used during conversation and the interacting parties kept more distance when not talking. This finding is consistent with those of Webb & Weber [14], and rather relevant given the prevalence of age-related hearing loss, which affects over half of the population aged 75 and above [2].

Our results also show that our target group is diverse and complex: various internal, highly personal variables strongly influence their interaction behaviour. Taking these differences into account is a major challenge for TERESA.

B. Observing the Elderly Interact via TERESA

We also conducted observational studies in which elderly visitors controlled TERESA while having a conversation with their peers. We found that the conversations were influenced by being mediated by a telepresence robot; many of the elderly were concentrating strongly on controlling the robot and as a result seemed to be less available for conversation. Or, as one participant remarked; "I can't do everything at the same time". Though this may have been a novelty effect, most of the conversations were about the robot. In addition, some of those interacting with the robot tended to give it orders (such as "follow me", "sit down"). They did however indicate feeling a presence; some even remarked that they saw no difference between talking with the robot or without it, describing it as much more 'present' than talking through a phone or Skype. Afterwards, many of the participants indicated that they enjoyed the experience.

These observations indicate that a telepresence robot could have added value, but that learning to control the robot is hard and requires much effort even after training, reducing the quality of the conversation. Introducing semi-autonomous navigation, as TERESA aims to do, could thus help make the robot more usable.

III. DETECTING & UNDERSTANDING EMOTIONAL REACTIONS

Making TERESA effective requires developing new algorithms that automate detection and tracking of human communicative cues. These include facial expressions, focus of visual attention, body gestures and postures, as well as speech-related acoustic features. In addition, it requires a new set of methods for the interpretation of human behavioural cues in terms of reactions to the robot shown by both the visitor and the interaction targets.

Fig. 2: Example of facial point tracking.

A. Automatic Understanding of the Visitor's Reactions

A key task is to build an automatic facial expression recognizer that can detect with high accuracy smiles, frowns, head nods/shakes and can estimate valence and arousal. We describe here the progress towards this goal, which includes implementation of a facial point tracker and a preliminary version of a facial expression recognizer.

1) Facial Point Tracking: We have built a facial point tracker to track the 2D and 3D location of the 66 facial landmarks as shown in Figure 2. In addition, the facial point tracker detects the face's 3D pose in terms of pitch, yaw and roll with respect to the camera's optical axis. The facial point tracker implements the Constrained Local Model (CLM) [4], [12]-based Discriminative Response Map Fitting (DRMF) method proposed in [1]. To improve processing speed, we developed a hybrid CPU-GPU implementation of the DRMF method, as described in [3]. By exploiting the potential data parallelism in the algorithm's Histogram-of-Gradient (HOG) feature extraction step and response map calculation step, this hybrid implementation is 5 times faster than the CPU-only baseline version [3], achieving real-time performance.

2) Preliminary Facial Expression Recogniser: Using the location of the tracked facial landmarks as input features, we have trained a multi-class support vector machine (SVM) classifier to perform facial expression recognition. In particular, the k-class SVM classifier consists of $k(k-1)/2$ binary SVM classifiers, each trained using examples from only two classes to find a hyperplane maximizing the margin between them. To classify unseen data, all binary classifiers are used and the overall decision is derived using majority vote. We have used Radial Basis Functions (RBFs) as the SVM kernel.

Our current implementation was trained on the Multi-PIE database [5]. Specifically, around 3500 images from subjects 1-170 have been used as training examples. The feature vector consists of the 3D location of the 66 facial landmarks tracked by the facial point tracker. Nonetheless, to eliminate unwanted influence of rigid motion and scaling, the faces are first registered to frontal pose before the feature vectors are calculated. Due to the content limitation of the Multi-PIE database, we have only trained a 5-class SVM capable of distinguishing between neutral, smile (happiness), scream (fear), surprise, and disgust. Nonetheless, this component can be easily extended to recognize new expressions by retraining the SVM when appropriate examples are available.

B. Automatic Pose Estimation of Interaction Targets

To respond socially to the people around it, TERESA must localise them and identify relevant features of their pose.

For this, we focus on the data from the Kinect depth sensor (see Section I-A), and evaluate its suitability for robot-based tracking of people in social interactions.

We collected a dataset, including ground truth, in a suitable context, which allows us to train and evaluate the person localisation algorithms. In line with the context of the project, we created a setup in which a visitor went through multiple cycles of approach, conversation and retreat with three interaction targets. Ground truth on the position and orientation of TERESA and the participants was acquired using an OptiTrack[TM] motion capture system.[2] 14 groups participated (56 participants), resulting in well over 6 hours of data.

Our experience with the Kinect skeletal tracker[3] on this data is that, after it has had a frame with a full (non-occluded) frontal view of a person, it is quite effective in tracking that person. The ratio between the Kinect skeletal tracker detecting at least one person and at least one person being in sensor range was 68.7%.

Our aims for now are to add more features, such as orientation of people and detecting who is part of a group. In addition, we aim to integrate information from face detection using the Dalsa camera mounted on the robot, to improve the performance and reliability of the localization module.

C. Automatic Understanding of Interaction Targets' Reactions

We have developed a fundamental speech emotion recognition system which identifies five basic categories, i.e., boredom, sadness, happiness, neutral, and anger. This system builds Hidden Markov Models for the classification by using acoustic features: f0, Mel-frequency Cepstrum Co-efficiency (MFCC), and energy. We trained the models using the SEMAINE corpus and the LDC Emotional Prosody Speech corpus. In addition, to resolve inter-speaker variations, we employed speaker adaptation techniques. This speech emotion recognition system produces labels, which might be possible subsets of features for higher levels of social states such as the quality of the conversation.

IV. SOCIAL NAVIGATION

TERESA is intended to operate among people in social environments. This poses several challenges related to robot navigation. Not only are obstacle and people avoidance required, but different goals than classical motion planning are usually pursued. For instance, in [13], the authors discuss path-planning problems for social robots, which include, among others, paths that minimize the interference with people (to avoid discomfort), or that maximize the probability of finding someone.

Motion planning requires the ability to predict the behaviour of the system, and a notion of costs (or utilities) associated to robot behaviours. Current motion planners for robot navigation cannot adequately perform social navigation because they do not attempt to predict human behaviour but

---

[2]www.naturalpoint.com/optitrack/

[3]msdn.microsoft.com/en-us/library/hh973074.aspx

instead just minimize travel time or path length, which does not in general translate to social navigation.

Social navigation in TERESA is novel in the following ways: first, we explicitly consider the need to predict human behaviours that may affect navigation with reasonable accuracy (see Section IV-A); Furthemore, our costs are related to social compliance, in order to set preferences among different motion behaviours. Instead of defining these cost functions by hand, TERESA encodes the social navigation behaviour as a set of costs learned from the controller using data from experiments. This allows the motion planners that are used to guide the TERESA system to adapt to different social contexts (see Section IV-B).

A. Environment & Human-Robot Interaction Models

Human navigation behaviours often depend on spatial variables. People move between rooms or offices following similar motion patterns; places of interests, like coffee machines, are relevant when considering where people are headed, etc. Ad-hoc modeling of these elements is time consuming and does not generalize to different environments. Machine learning techniques can be used to infer these models from data. Then, they can be used by the navigation planners in a predictive fashion to improve the behaviour of the robot in social navigation environments, by considering human intentions when planning.

We analysed the use of spatial Hidden Markov Models to learn human trajectories and goals from observations. To provide a spatial modeling framework for TERESA, we improved the original HMM framework to enable online learning from partial and noisy human trajectories without special knowledge about the trajectory goals. The system is able to automatically discover points of interest by analyzing the directed graph of observed human trajectories, the probability of reaching those goals, and the motion of people in the environment.

Initial low-scale tests have been performed with TERESA in the coffee/snack machine area of one of the buildings at University Pablo de Olavide. This is an area of 4,30 x 11,80 meters accessible from two corridors and containing several points of interest such as coffee and snack machines, bathroom doors, and a water fountain.

Our system generates a topological and metrical network of positions and determines the points of interest in the area. Fig. 3 shows a sequence of frames running the prediction service for a trajectory at a time horizon of 3 seconds. Qualitatively, the system is able to predict coherent goals and positions.

We have also benchmarked the current algorithm using the Edinburgh Informatics Forum Pedestrian Database [10]. The data covers several months of observations which has resulted in about 92000 observed trajectories. This dataset was used for validation, since we know *a priori* the endpoint for each trajectory. After generating the model, two different metrics were applied in the prediction phase for 1000 test trajectories, at 25%, 50% and 75% of the full length of each trajectory. Table I shows some initial results. The
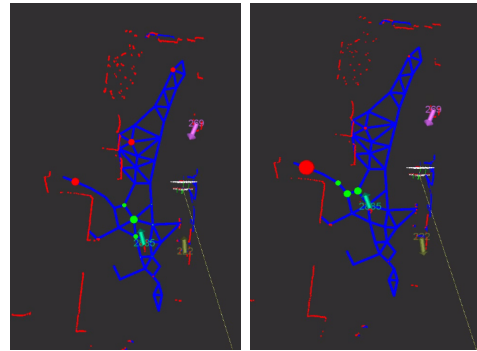

Fig. 3: *Running the prediction service for the coffee machine area at UPO at two timesteps of the same trajectory for a time horizon of 3 seconds. The arrow represents the current position of the tracked person; the red dots mark the position of the possible goals; the size of each dot is directly related to the probability of reaching the goal. The position and size of the green dots represent the probability distribution of the next areas in the specified time horizon.*

number of discovered goals in the scenario are 12, which means a uniform probability of 0.08 for each goal if nothing else is known. It can be seen how the system enhances the predictions capabilities. As expected, the prediction power increases as new information is received.

Trajectory time	Metric 1	Metric 2
25%	0.139279	4.12487
50%	0.189846	3.50888
75%	0.245137	2.85467

TABLE I: *Experimental validation using the Edinburgh data set. Metric 1 is the mean estimated probability for the ground truth goal. Metric 2 is the main error on the position of the goal.*

B. Social Path Planning Under Uncertainty

We are developing a planner for efficient social navigation based on Rapidly-Exploring Random Trees (RRTs) [8]. We use RRTs as a baseline as they allow exploring the navigation state space fast and take nonholonomic and kynodynamic constraints into account. We are extending the planners by adding the social navigation conventions previously learned from data.

Specifically, we need to consider within the configuration space all the relevant features for social navigation, like relative people distances and orientations. Furthermore, we have analysed different ways of incorporating the social cost functions learned in Section V. Some extensions of RRTs involve cost spaces, such as Risk-RRT, which includes costs based on proxemics [11] as well as RRT* [8] or Transition-based RRT [7]. An initial path planner has been integrated with a classical obstacle avoidance controller for path execution and ported to TERESA. The main challenge is to integrate the predictions about people's poses derived from the models from the previous section. Furthermore, it is complicated to encode into a single cost function all the social conventions, and thus the planner will need to adapt the cost function employed depending on the social situation. Finally, another line of work in the project is analyzing ways to consider these models within the cost learning stage described in the next section.

V. Cost Functions

Achieving behaviour that respects social norms is a key challenge in robotics because social behaviour, being so complex, cannot be easily hard-coded into the cost function that underlies the robot's decision making.

Instead, TERESA aims to learn these cost functions from data by associating the features extracted by the modules described in Section III with different sources of feedback collected in natural interaction scenarios.

Figure 4 shows the learning pipeline for the project. For each conversation and navigation scenario, data (features and feedback) are collected and fed into different learning algorithms. This produces several cost functions that are then integrated and act as input to a planner that selects appropriate actions.
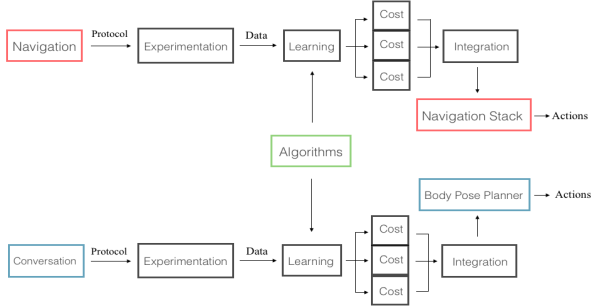

Fig. 4: Learning and planning pipeline for TERESA.

A. Experimental Setup

Since datasets of robots in social situations are not readily available, performing successful experimentation in real life scenarios is one of the great challenges and contributions of the project.

To generate data for learning cost functions, we conducted an experiment with human subjects. The experiment involves two rooms: the *interaction room* and the *visitor room*. In the visitor room a visitor observes, through the telepresence robot, what is going on in the interaction room. Also present in the visitor room but hidden from the visitor is a *Wizard of Oz* (WoZ). Although the visitor believes the robot is autonomous, it is actually controlled by the WoZ. In the interaction room there is TERESA, an interaction target, and several confederates. A schematic of the general setup is shown in Figure 5. Apart from recording from all robot sensors, we also record several feedback signals, namely:

- *Expression dataset* : Implicit feedback from the visitors facial expressions as described in Section III;
- *Direct feedback dataset*: Direct instantaneous visitor feedback through button presses whenever unacceptable robot behaviour was observed;
- *Expert quality dataset*: Global visitor feedback, rating the general behaviour of the robot from 1 (unacceptable) to 5 (socially perfect);
- *Expert dataset*: Expert demonstrations of behaviour as prescribed by the Wizard of Oz.

The Wizard of Oz methodology [9] is key because it generates realistic feedback about how subjects react to an autonomous robot before such a robot has been created.
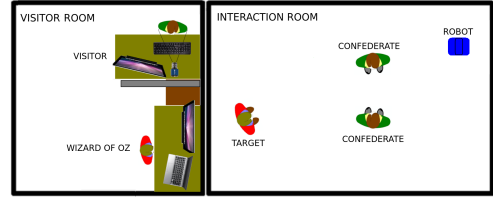

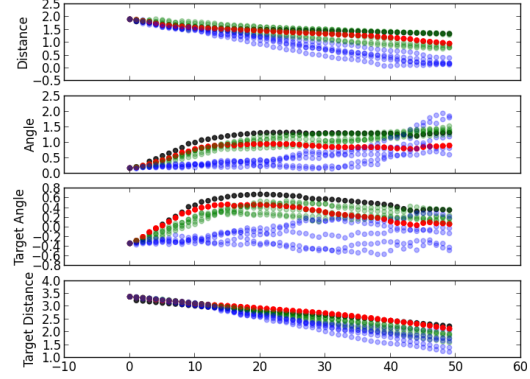*Fig. 5: The general Wizard of Oz Setup.*

Fig. 6: Learning using IRL. Expert path (red-black) in time-steps (x axis), Before learning (blue), After Learning (green).

Using this setup, we collected a total of 10 hours of sensor data from 20 experiments and more than 25 subjects, amounting to 720 trajectories and 124Gb of Data.

B. Learning Cost Functions

Using data from our navigation experiments, we produced and evaluated two cost functions.

a) Inverse Reinforcement Learning (IRL): The first cost function exploits the fact that the WoZ can be considered an expert at the navigation task whom TERESA can emulate. To accomplish this, we employ Maximum Causal Entropy Inverse Reinforcement Learning [15]. Qualitative results from this work, shown in Figure 6, demonstrate that after learning the robot is capable of imitating the expert if it is given the same initial conditions. Furthermore, Figure 7 shows the average Euclidean distance between robot and expert trajectories in a cross validation dataset for different learning features and methods from which our specific IRL feature set performs the best, outperforming a policy learned by imitation, i.e. supervised learning.

b) Direct Feedback Learning: Another source of feedback from the experiments came from the visitor pressing a button whenever TERESA performed socially inappropriate actions (the direct feedback dataset).

Treating these presses as 'labels' of bad behaviour, we used supervised learning methods to distinguish between neutral states and bad states. Considering the sparsity of labels in the datasets, we devised a metric that favours classifiers that predict neutral states as bad states (false positives) rather than the opposite. Using this metric, we performed a grid search over possible parameters of different classifiers. This resulted in choosing a support vector machine with biased class weights for this cost function.
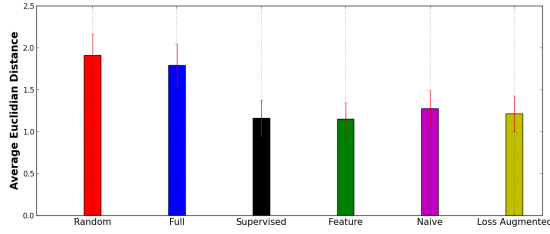
Fig. 7: *Euclidean distance from cross validation examples for different learned policies.*

C. Integration

Another key challenge is to integrate the different cost functions we obtain from each feedback signal. Poor integration can bring about arbitrary behaviour so special care must be taken. We considered two methods of integrating the cost functions, a *naive* approach and a *loss augmentation* approach. While the naive method simply constructs a linear combination of the two reward functions, loss augmentation uses the second (supervised) cost function within the IRL algorithm, making sure that the resulting cost function is consistent with the expert while at the same time taking direct user feedback into account. The performance of the two methods was evaluated using the metric utilised in assessing the performance of IRL (direct comparison), and by counting the number of *intrusion states* that the policies resulting from these cost functions visit (indirect comparison). These intrusion states were in turn derived from the direct feedback data. Figure 7 (magenta, yellow) and Table II show that the loss augmented integration performs better in both metrics.

Policy	Intrusions μ ($\pm\sigma$)
Data	11.19 ($\pm$13.2)
Feature	32.57($\pm$25.6)
Imitation Learning	12.2($\pm$18.1)
Naive	14.8($\pm$18.1)
Loss Augmented	11.2($\pm$16.0)

TABLE II: *Indirect comparison of cost functions in terms of intrusions, i.e., the average number of "bad" states encountered during a trajectory towards a target, as defined by the information in the direct feedback dataset.*

VI. Conclusions and Future Directions

By the end of the project, TERESA will display robust semi-autonomous, socially intelligent behaviour in a challenging real-world environment. In terms of perception, it will be able to recognize social signals by combining various modalities such as sounds, images and depth sensors, with state-of-art methods. Having access to these multimodal social features will allow the robot to act in a socially appropriate manner based on its embedded social cost functions. In this way, TERESA will make visitors feel and seem more present in the physical space in which the robot is acting.

We envision a scenario in which the visitor can easily instruct the robot to approach a person, or a group of people in order to engage in conversation. While navigating, the robot respects social norms, operating in a safe and non-intrusive manner, by maintaing an appropriate velocity, distance and orientation to the surrounding people. While engaged in conversation, TERESA reacts autonomously to social cues. For example, if TERESA detects that either of the conversation participants has hearing problems, it autonomously adjusts its body pose to improve communication; it controls its height and head tilt when a person sits down during a conversation, in order to maintain visual contact; and if the interaction target is performing other activities, it tracks them and adjusts its body pose accordingly so as to allow seamless communicatiom.

The progress described in this paper is only a first step towards what the TERESA project aims to contribute to a variety of fields such as robotic perception, navigation and learning, social signal processing and human-robot interaction, making TERESA a multidisciplinary effort that will advance the state of the art in each of these domains.

References

[1] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *IEEE CVPR*, pages 3444–3451, 2013.

[2] N. Blevins, D. Deschler, and L. Park. Presbycusis. www.uptodate.com/contents/presbycusis, 2015. Online; accessed 7 March 2015.

[3] S. Cheng, A. Asthana, S. Zafeiriou, J. Shen, and M. Pantic. Real-time generic face tracking in the wild with cuda. In *Proc. ACM Multimedia Systems Conference*, pages 148–151, 2014.

[4] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, volume 3, pages 929–938, 2006.

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[6] E. T. Hall. *The Hidden Dimension*. Anchor Books New York, 1966.

[7] L. Jaillet, J. Cortés, and T. Siméon. Transition-based RRT for path planning in continuous cost spaces. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Nice, France, Sept. 2008.

[8] S. Karaman and E. Frazzoli. Sampling-based algorithms for optimal motion planning. *CoRR*, abs/1105.1186, 2011.

[9] W. B. Knox, S. Spaulding, and C. Breazeal. Learning social interaction from the wizard: A proposal. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[10] B. Majecka. Statistical models of pedestrian behaviour in the forum. http://homepages.inf.ed.ac.uk/rbf/FORUMTRACKING/, 2009. MSc Dissertation, School of Informatics, University of Edinburgh.

[11] J. Rios-Martinez, C. Laugier, and A. Spalanzani. Understanding human interaction for probabilistic autonomous navigation using Risk-RRT approach. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, United States, Sept. 2011.

[12] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011.

[13] G. D. Tipaldi and K. O. Arras. Planning problems for social robots. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS'11)*, Freiburg, Germany, 2011.

[14] J. D. Webb and M. J. Weber. Influence of sensory abilities on the interpersonal distance of the elderly. *Environment and behavior*, 35(5):695–711, 2003.

[15] B. D. Ziebart, J. A. Bagnell, and A. K. Dey. Modeling interaction via the principle of maximum causal entropy. 2010.