Problem and context. Commercially available telecommunication systems do not provide the sensation of self-presence, for users do not get the feeling of being spatially present at the remote site, nor do the people they interact with feel them as physically present (a concept called “teleexistence” [1]). The most recent prototypes give indeed the user the feeling of being inside a robot via haptic feedback - projecting the person’s image onto the remote robot’s reflective surface provides their remote partners with the impression that they are really present [1]. Realism in robot surrogates or “avatars”, however, is not just about building machines which *look* like the people controlling them: as the way we express emotions is such an important component of who we are, enabling robots to imitate this characteristics is a crucial element of any realistic surrogate [2]. Electroencephalography (EEG) provides a means to detect and recognise brain electrical signals associated with various cognitive functions [4], including emotions [5,6,7].

Goal: *We seek to develop an ‘emotional’ robotic avatar system (Figure 1) in which the affective state of an individual is detected and recognised from EEG signals captured in real-time in their natural environment, and expressed by the robot surrogate through appropriate facial expressions and bodily gestures*.

Significance and timeliness. Effective emotional avatar systems do not only have the potential to significantly enhance the experience of remote personal presence, but can also bring humanoid robots closer to having a natural interaction with people (e.g. as surrogate teachers working with schoolchildren). In the longer term, they could provide realistic body surrogates for people with disabilities so allowing a paraplegic person, for instance, to enjoy a degree of physical interaction with others. In the UK alone over 11 million people have a limiting long-term illness or disability (source: Family Resource Survey). Commercially, the robotics telepresence market is expected generate over $136.90 million by 2022 (Research&Markets), while the collaborative robotics sector will increase tenfold by 2020 (IEEE Spectrum).

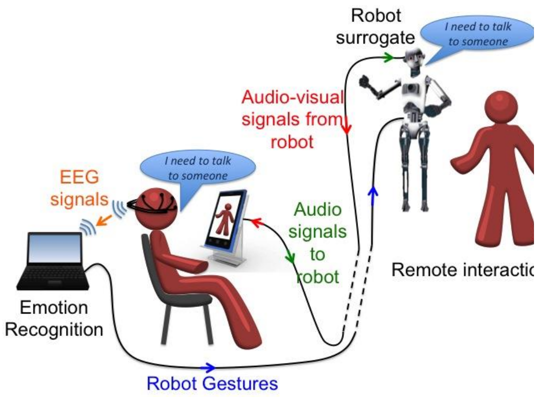

Figure 1. *AVATAR concept*.

Our goal is ambitious, as the recognition of people’s affective states based on brain signals has been tested to date mainly in controlled settings. Real-time recognition and control are still beyond the capabilities of current systems. Nevertheless, in our professional view, recent advances in social robotics, brain signal analysis and machine learning, together with the availability of cheap but reliable electroencephalographic devices and affordable, easy-to-control humanoid robots make the creation of a robot emotional avatar now possible [3]. Solving the technical challenges involved will lead to further significant progress in electroencephalographic analysis and to the development of novel machine learning methodologies which will build on and reinforce current trends in robust artificial intelligence.

State-of-the-art. Emotion classification from EEG data remains relatively unexplored [5,6,7]. Typically, a small set of emotions is sought to be detected from signals obtained from electrodes located in the areas of the scalp where the associated brain activities are believed to reside. The emotions’ lateral location in the frontal region of the brain is also considered [8]. Emotions are conventionally represented using the cognitive Valence/Arousal model [6,7], where affective states are labelled as positive vs negative, calm vs excited, although models which tie them to cognition exist. Support Vector Machines or Linear Discriminant Analysis are then typically applied for classification purposes [5,8].

Most relevantly, brain patterns are heavily affected by external stimuli such as background noise, or the presence of other people moving and talking in the same area. The subject’s own eye-blinking reflex and unconscious body movements are also constituent to brain activity patterns. Such ‘distractors’ have been so far addressed by restricting the recognition to artificial, controlled environments [9,10]. High- or low-pass filters are used to dismiss perturbations in brain activity associated with body movements [12]. ‘Electrooculograms’ [11] are applied to get rid of corrupted data segments associated with eye blinking. In addition, participants are required to remain still with their eyes either open or closed during EEG recording sessions to mitigate the unwanted influence of muscular movements [9]. Emotion elicitation procedures are mostly based on audio-visual stimuli, including music [13], videos [14], images or facial expressions, and also assume clean and controlled environments. Individuals whose emotions are to be classified are shown temporally spaced projections of multiple EEG recordings (known as ‘trials’) for each emotional state.

Few limited efforts have been made to move from offline to real-time, 'online emotion' recognition in unconstrained environments [15]. Jatupaiboon et al [16] have recently published details of a real-time EEG system for differentiating between just two classes: happy and unhappy, while [17] developed a system capable of recognising four emotions (pleasant, happy, frightened & angry). While these approaches have shown some promise in terms of real-time emotion recognition, they remain constrained to a well-defined and restricted operating environment. The reliable detection of a user's emotional state through single-trial EEG data in real environments remains a significant challenge, one which this proposal seeks to address.

Our approach to the creation of effective robotic surrogates follows recent studies which have shown that a robot's gestures, poses or facial expressions have a positive effect on how people perceive them [18,19]. Co-Investigator Crook has recently shown that robots mirroring the body pose of their human partners are perceived as more human-like [CI1]. Controlling the robot to perform movements or expressions which convey the detected emotions constitutes another major research question.

The **objective of this research** is the development of a robotic emotional avatar system able to:

(1) recognise the user's emotion (affective state) in their natural environment, despite multiple distractors;
(2) perform recognition in real time (on a human perception scale), as EEG signals have to be segmented and recognised moment by moment, so that the robot may enact the recognised emotions immediately;
(3) detect in time and spatially localise, across the 32 electrodes, EEG patterns that encode the sought emotions for all brain areas may potentially contain information pertaining affective states;
(4) use the recognised emotion to generate, via an appropriate control strategy, command signals which allow the robot to express such emotions via gestures and facial expressions.

The **material outcomes** of this project will be:

(a) a repository of efficient software implementing electroencephalographic analysis, real-time brain signal classification and robot control, which will be made available open source;
(b) a new extensive dataset of brain signals acquired under various environmental conditions to simulate real world situations, completed with human-generated annotations, made available to all researchers;
(c) a working prototype of real-time emotional robotic avatar, whose elements are detailed in Methodology.

Equipment. The three groups led by Cuzzolin, Crook and Camilleri are already in possession of state-of-the-art equipment required to implement the various elements of AVATAR. Facilities at the Brookes Cognitive Robotics Lab (Crook) currently include a Baxter robot (http://www.rethinkrobotics.com/baxter/), two NAO units (https://www.aldebaran.com/en/cool-robots/nao) and the Robothespian robot "Artie" (https://www.youtube.com/watch?v=i5dPafSxr9g), the latter most suitable to the task at hand. Equipment is constantly upgraded. The AI and Vision group (Cuzzolin) owns a significant machine learning code base in C#, Python and Matlab. Malta (Camilleri) possesses a 32-channel g.tec, an 8-channel wireless Enobio and two 14-channel wireless Emotiv EEG acquisition systems. Our approach aims to be cross-platform, as it can be implemented on both relatively cheap (e.g. Emotiv, NAO) and high-end equipment (Artie, G.TEC.).

Methodology. The avatar system will be composed of (Figure 2): Stage 1 - EEG signal acquisition, including an emotion elicitation procedure; Stage 2 - feature extraction from the EEG signals; Stage 3 - detection (in time) and localisation (in terms of groups of electrodes) of relevant EEG patterns; Stage 4 - robust classification of the detected segments; Stage 5 - robot control strategy to reproduce the detected emotions. In addition, a Stage 0 concerning the study of the relation between affective states and gesture, poses or expressions and a Stage 6 concerning our testing and validation strategy will be implemented.

Stage 0 [Leader: Crook] – The **relationship between poses/gestures and emotions** will be studied throughout the entire project, under the advice of Dr Michael Pilling and Dr Sanjay Kumar from Oxford Brookes' Psychology department. We will classify emotions into: afraid, angry, disgusted, happy, sad, surprised and 'neutral', categories which can be directly mapped to unique intervals in the Arousal/Valence scale, and are the primary basis for classification of facial expressions in existing emotion databases (e.g. KDEF [20], EmotiW [21], MMI [22]). Psychological research has shown that emotions can be differentiated both by characteristic patterns of body movements and by their static postures. Such cues are effective even when viewed from a distance or when the face is not visible [42]. Studies on the perception of emotion in terms of body movements have shown that affective states can be reliably inferred from these motion cues alone [43]. A strong link between emotion expression and gesture production has been suggested where arousal and potency are found to be strong predictors of gesture dynamics [44]. Evidence also suggests that such patterns of body movement are processed rapidly and automatically in the brain, and that they influence our perception of facial expressions and tone of voice [45]. Nonverbal cues such as head and eyes position, mouth state, neck rotation, body pose and gestures exert great influence in shaping emotional perception, which is reflected in greater activity in the dorsolateral prefrontal cortex [46].

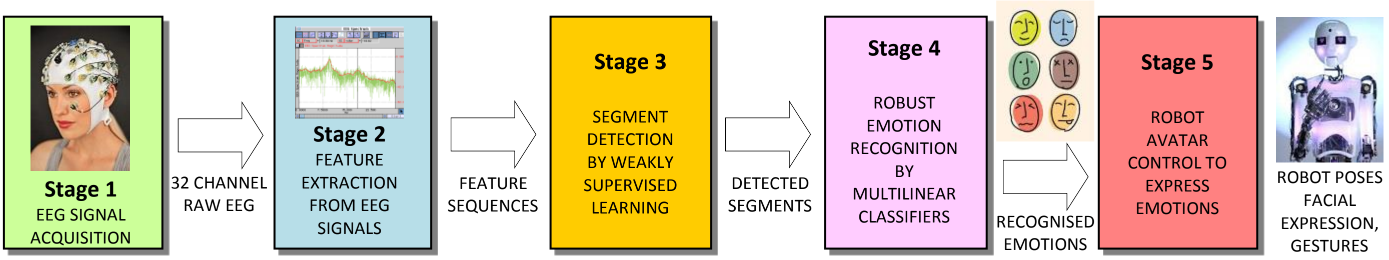
Stage 1
EEG SIGNAL ACQUISITION
32 CHANNEL RAW EEG
Stage 2
FEATURE EXTRACTION FROM EEG SIGNALS
FEATURE SEQUENCES
Stage 3
SEGMENT DETECTION BY WEAKLY SUPERVISED LEARNING
DETECTED SEGMENTS
Stage 4
ROBUST EMOTION RECOGNITION BY MULTILINEAR CLASSIFIERS
RECOGNISED EMOTIONS
Stage 5
ROBOT AVATAR CONTROL TO EXPRESS EMOTIONS
ROBOT POSES FACIAL EXPRESSION, GESTURES

Figure 2. *Block diagram of the elements of the robotic avatar system. Stages 0 and 6 not included.*

Stage 1 [Leader: Camilleri] – EEG signal acquisition. Emotion elicitation experiments are carried out by measuring the emotional response of subjects to pictures, as in the International Affective Picture System (IAPS), and to sounds (cfr. the International Affective Digitised Sounds (IADS) system). Other emotion elicitation protocols have also been developed around music video clips [23]. Here we shall preferably adopt a database of calibrated stimuli, such as the IAPS or IADS. These will be used to elicit a spontaneous emotion in the subject, while measuring a full-scalp 32-electrode EEG, and having the subjects rate the picture, sound or video in terms of arousal and valence. Participants for this study will be normal healthy adult volunteers (typical exclusion criteria such as a history of mental health conditions, recent intake of medicine, drug abuse will apply), recruited through notices, adverts and invitations. The number of subjects recruited is constrained by the duration of the acquisition which requires a substantial time to assemble the scalp cap and electrodes, perform the acquisition protocol and subsequently dismantle the electrodes and clean them up. This study will aim to acquire data from at least 20 subjects, which is comparable to that found in the literature [24]. Camilleri's team have extensive experience in recruiting volunteers for such studies, designing EEG data acquisition protocols, and processing EEG data.

Stage 2 [Leader: Camilleri; Cuzzolin] – The **extraction of feature measurements from EEG signals** is crucial for emotion discrimination. We will explore and refine two complementary approaches.

The EEG signal is a multi-channel signal obtained from several scalp electrodes that pick up the electrical activity from the brain as well as from other sources in the body (muscle activations, eye blinks and eye movements). Pre-processing for the reduction of such artifacts will then need to be carried out [25,26]. Because of the nature of EEG signals, multi-channel analysis techniques [27] (of which Camilleri's team have extensive experience [CO1,CO4]) will be adopted, in combination with an original spatial filtering algorithm [CO5,CO6] which enhance classification performance where phase plays an important part in the discrimination. Within EEG-based emotion recognition spatial filtering has only been tested on music-induced emotions so far [28]. 'Univariate' analysis of individual EEG channels [6] will be tested first to build a baseline for comparison. Subsequently the team will investigate the use of unsupervised Independent Component Analysis (ICA) to study the source components involved in the EEG signal during emotion elicitation, and their exploitation for emotion recognition. Next, spatial filtering techniques based around Common Spatial Patterns (CSP) [CO11] will be applied to EEG emotion analysis and recognition. Camilleri has recently generalised the CSP method to a complex variant, known as the Analytic CSP (ACSP) [CO8], which affords the explicit handling of the phase of the signal alongside its magnitude characteristics.

Recent developments have seen Neural Networks (NNs) with an increased depth of network layers and more effective learning strategies back to the forefront of machine learning [29], as they allow us to learn multiple levels of feature representation at different abstraction levels. These 'deep' networks have recently achieved state-of-the-art performance on a variety of classification problems (e.g. object detection [30], action recognition [31], action localisation [32]). Cuzzolin's team recently developed a novel pipeline for online action recognition and localisation based on deep learned features [PI9] which outperformed the state of the art on the most recent and challenging benchmarks. The suitability of deep learning has started to be explored for brain-computer interfacing [33] and for EEG sleep analysis [34] too. In this project we propose therefore to explore, in parallel, various deep learning approaches to EEG signal encoding which, rather than extracting hand-crafted features from them, present raw signal values or early-stage pre-processed signals as input to an appropriate deep neural network to generate discriminative features.

Stage 3 [Leader: Cuzzolin] – Localisation of relevant EEG segments via novel weakly-supervised classification approaches. Given a labelled training set of "bags" (here, EEG datastreams), Multiple Instance Learning [35] looks at all EEG subsequences (for each group of channels and each time interval), called 'instances', to build a SVM model for each class label. After an iterative process, only the most discriminative instances in each positive 'bag' (i.e., EEG datastreams displaying an emotion) are retained and used to learn a SVM classifier, which can later be applied to new datastreams to locate and classify segments and channels in which emotions are displayed. MIL's successful application to human action recognition was recently brought forward by Cuzzolin and his group [PI5,PI6,PI8]. A competing approach to weakly supervised classification by Siva et al. [36] has shown excellent results by using far off negative examples to describe the data's inter-class structure ('negative mining'). The CRANE algorithm [37] has generated further improvements by using negative examples merely to attribute 'penalties' to uncertain positive instances. CRANE is only one of a family of methods which exploit distances between weakly labelled instances. We will thus explore ways of mining information from the whole training distance matrix to achieve more robust assessments. The performance of the developed emotion detector from EEG will be validated on the collected data (Stage 1), by splitting the dataset into training, validation and testing folds. Results will be fed to Stage 4.

Stage 4 [Leader: Cuzzolin] – Robust classification of detected segments via novel multilinear classifiers. Early bilinear classifiers [38] have been proposed to model the influence of a single 'style' variable [PI1]. They allow, for instance, to build a classifier which given an EEG signal captured under an environmental condition not seen in the training set (e.g. presence of noise), can estimate both distractor and emotion parameters, improving recognition performance. We will start by testing the ability of bilinear models to cope with single distractors, and their individual effect on performance. EEG signals, however, are affected by many such nuisance factors. Methods such as High-Order SVD [39] have been proposed to tackle their influence by decomposing a tensor into its constituent factors. Efforts have mainly focussed on the generalization of unsupervised dimensionality reduction to tensors (MPCA) [40], by mapping tensors to lower dimensional ones of the same order and extract features from their "core" tensor.

The PI [PI2] has instead recently proposed the development of true 'multilinear' classifiers [PI3,PI4] able to model multiple nuisance factors and alternatively estimate nuisance (e.g. environmental noise) and 'content' (emotional) variables by minimising the least square reconstruction error on the training data. A set of style-specific maps are learned by HOSVD of the training set. When a new observation in a combination of styles not in the training set is presented, Expectation-Maximisation is applied to alternatively learn a new style matrix and classify the content of the observation.

Current bilinear or multilinear frameworks, however, classify test observations in batches rather than on an individual basis. A crucial step of the proposed research will be the development of an online (incremental) version of these classifiers to tackle real-time requirements. We will as well develop an incremental training setting in which the multilinear interaction model is efficiently updated whenever new observations become available, to allow the system to continuously learn from experience.

Stage 5 [Leader: Crook] – Robot control strategy to express recognised emotions. We will employ the commercially available RoboThespian [41] (which appears in Figure 1), a humanoid robot with fully configurable torso, arms, head, eyes and fingers. The robot's features can be controlled via a standard XML interface to increase the realism of the selected expression (Stage 0). Visual feedback will also be provided to the person coupled with the robot via an iPad for a complete immersive experience.

The key challenge here is to map classes of affective states to control sequences in the robot avatar that will enable it to mimic these emotions through the appropriate use of head and eyes position, face colour, mouth state, neck rotation, body pose and gestures. The mapping will be achieved by defining a set of key robot poses for each emotional state, together with the ability to generate random variations on these poses that are in character with the emotion. Movements need to be expressed as changes in the joint angles of the robot's kinematic model. Parameters associated with each emotion will define the characteristics of movements and poses related to that emotion. For example, anger can be expressed with high amplitude fast limb movements, while sadness with low amplitude slow transitions between poses.

Stage 0 will provide the key robot poses and movement parameters that are associated with each class of emotion. Stage 5 will create the software modules that will use these to control the robot so that it can mimic the corresponding emotions. A key principle here is to develop software modules that can be used to generate emotional movements on a range of robotic platforms beyond the one we are using on this project. To this end, we will implement these modules using the Robot Operating System (ROS), which is widely recognised as the international standard for developing robot software.

Stage 6 [All investigators] – Validation and prototype

The core EEG analysis components (Stages 2,3 and 4) will first be validated on BCI data already in our possession, designed to test multiple nuisance factors. Camilleri's team has developed a BCI music-player application based upon a neurophysiological phenomenon known as Steady State Visually Evoked Potential [CO6] which allows a subject to attend to one of a number of flickering visual stimuli, each associated with a specific music-player command. Camilleri and Cuzzolin have recently acquired EEG data of subjects using this BCI application while they were immersed in visual and audio distractors and while in motion, contaminating the data with nuisance signals. Next, emotionally-elicited EEG signals recently made publicly available (DEAP dataset [23]) will be analysed. Success will be measured in terms of commonly accepted classification performance measures such as precision, recall and F1 score. Later on a novel benchmark on EEG emotion recognition in unconstrained settings will be collected to test the overall prototype, including various feature extractors (Stage 2) and localisation approaches (Stage 3). Human testers will assess the expressivity of the robot's enactment of the recognised emotions (Stage 5). Annotated data and code will be made available open-source to all researchers in this and related areas.

Project management. Day-to-day work will be conducted by two Research Assistants (RA1 and RA2) and a Local Researcher (LR) in Malta. The LR will organise data collection, form user groups and analyse different EEG feature extraction techniques (Stages 1 and 2). RA1 (at Oxford Brookes University under Cuzzolin's supervision) will mainly focus on the machine learning-intensive Stages 3 and 4. RA2 (at Oxford Brookes under Crook's supervision) will work on the psychological analysis (Stage 0) and the robot control elements of Stage 5. Regular monthly meetings including PI, CIs, RAs, and the LR will coordinate the work of the three subgroups. Regular exchange visits between Oxford and Malta will take place and have been duly costed. An advisory board has been set up to help steering the project towards success. Dr Michael Pilling and Dr Sanjay Kuman of Brookes' psychology department will advise us on the study conducted in Stage 0 and the human assessment of the expressivity of the prototype avatar (Stage 6). Dr Matthias Rolf, Senior Lecturer in robotics at Brookes, will assist the team on the robot control strategy (Stage 5).

Stage/Months	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24	25-27	28-30	31-33	34-36
Stage 0	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
Stage 1	🟩	🟩	🟩	🟩								
Stage 2	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦				
Stage 3			🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧		
Stage 4					🟪	🟪	🟪	🟪	🟪	🟪	🟪	🟪
Stage 5							🟥	🟥	🟥	🟥	🟥	🟥
Stage 6					🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

Figure 3 – Workplan. *A Gantt chart of the project's work programme (36 months).*

Publication and dissemination. Results will be published at top-tier conferences and journals in machine learning, bioengineering and robotics, starting from end of Year 1. NIPS and ICML are the world's most important Machine Learning venues, with acceptance rates of 15% for posters and 3-5% for orals. Oral papers are presented in front of an audience of 1200 people, among which all the best researchers in the field and representatives from companies such as Microsoft and Google. Starting from the end of year we will target publication on top journals such as IEEE Pattern Analysis and Machine Intelligence (impact factor 5.781). We will also target the brain signal processing, affective computing and biomedical engineering audience through top journals such as the IEEE Tr. on Biomedical Engineering and the IEEE Tr. on Neural Systems and Rehabilitation Engineering, and conferences such as the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC) and the IEEE Conference on Neural Engineering (IEEE NER). We will aim for two or three conference publications and one or two journals each year. A web site dedicated to the project will be built to disseminate the results to the academic audience, attract business partners with the goal of setting up additional patents or industrial partnerships, make the datasets gathered in the course of the research publicly available, and contact new partners to kick-start follow-up projects at both national and European level.