Engage/Disengage: Control Triggers for Immersive Virtual and Robotic Avatars

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

Teleoperation encompasses the use of software and hardware interfaces to remotely control mechanical devices. In such teleoperational systems, the success of the task often relies on the factors of high-fidelity command/responses, low latency, and accurate or high-resolution feedback. Ideally, a telepresence humanoid robotic system, which faithfully replicates the operatorâĂŹs gesture and accurately relays sensory feedback to the operator, should achieve tele-embodiment and reciprocally enable an immersive sense of presence for the remote operator. In this paper, we propose and explore the feasibility of an EMG-based control interface for disengagement from a fully immersive humanoid telepresence robot. Our control interface enables a human operator to disengage from an embodied robotic or virtual avatar - where the operator is manually and verbally constrained by the avatarâĂŹs replication of the operatorâĂŹs gestures. Our system makes use of a support vector machine (SVM) classifier, with 94

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords

Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

INTRODUCTION

The field of telepresence has the goal of enabling an operator to feel like they are in a remote environment through the use of various technologies [19,22,35,46]. Fully achieving this feat can amount to a reduction in the time-cost of physical travel, and would open new opportunities for those with limited mobility - be it to constraints related to health or age. At its current state, commercially available telepresence systems focus on two parts: visual feedback [11] and mobility [23,40]. Systems such as the VGo^{TM} telepresence robot are in essence a video display/camera on a stick that is mounted on wheel-based platform [54]. Others such as the Double^{TM} [53] telepresence robot or the AMY^{TM} telepresence robot have similar mobile structures, with the varying difference in the differential drive - inverted pendulum mechanism. Seemingly, these commercial systems are designed for the use case of a remote office worker or even a remote student wanting to attend class. In the remote site, a software application allows the operator to control the movement and direction of the robot while video streaming the local environment. In parallel, the robot's video display video streams the remote userâĂŹs environment - creating and auditory and visual feedback loop that allows the operator to engage remotely. More simplistic approaches to telepresence are found in teleconferences systems that project a remote attendee across a boardroom. Inherently, to fully-achieve immersive telepresence factors such sensory feedback (auditory, visual, tactile), latency, mobility and other factors involving the manipulation of the remote environment must be considered and are actively being researched [20,47]. Exploring telepresence even further, we arrive at the notion of tele-embodiment - with the overarching goal of enabling avatars to represent a remote human in the real world or even in a virtual world - in essence, surrogates. Such avatars, be it as virtual representations or anthropomorphic robots would be required to faithfully replicate the operator's gesture and in parallel accurately interpret and relay any feedback (i.e. visual, auditory, tactile) to the operator. The latter is crucial, for in reality our actions change the world as much as its feedback changes us. We can think of this in terms of the sense-think-act cycle proposed in traditional AI and used to functionally decompose the tasks of robots. But, the exploration of tele-embodiment brings forth many challenges beyond those considered in telepresence. Social questions arise [37,38,49], questions regarding appropriate control interfaces, VR sickness, and questions regarding feedback representation are being researched. In this paper, we focus on the problem of disengaging from such embodied avatar; where a fully-immersed operator is manually and verbally constrained by the avatar's real-time replication of the operator's gestures. To this end, we propose an EMG-based system that accurately detects tongue gestures which are interpreted by our robotic system as signals to disengage. The purpose of our EMG-based âĂIJdisengagementâĂİ system is similar to that of assistive technologies that provide disabled individuals with alternative means of communication or control. For example, using eye gaze and eye movement to transcribe words, and the use of tongue gestures as a control

CHI'16, May 07–12, 2016, San Jose, CA, USA

ISBN 123-4567-24-567/08/06...$15.00

DOI: http://dx.doi.org/10.475/123_4

interface for mechanical systems such as electric wheelchairs [12,29]

STATE OF THE ART

The study at hand relates to the area of tele-robotics as well as control interfaces used in robotics. We use the terms "local" and "remote" to refer to the environment where avatar and the operator are located respectively. Immersion and Tele-Embodiment In everyday human interaction, non-verbal cues such as postures and gestures play an important role in relaying contextual and situational information [4]. Commercially available telepresence products go an extra step beyond video conferencing and enable remote users with mobility. Studies show that these systems increase the local participantsâĂŹ sense of presence of the remote user, and lead to a reciprocal sense of presence for the remote user [4,7]. However, the addition of a physical embodiment (mobile robot) to a video display showing a live feed of a remote user could yield to what is known as "dual ecologies" [25]. This creates a sort of dissonant dual reality for the local participant: a local reality where the robotic avatar is situated and a reality where the operator is located remotely. The problem of dual ecologies arises when the verbal intentions of the remote operator does not match the bodily projections of the robot (i.e. the gestures made when asking a question) [24,26]. Although mobility increases the sense of presence, it is known that the transmission of information in face-to-face communication combines the temporal and spatial features of boily gestures and speech. Thus, the use of such contemporary telepresence systems creates a "fragmented and relatively ineffective" relationship with the features or gestures that are symmetric to verbal communication [16,18]. In essence, it handicaps the remote participant from fully expressing and conveying a message, and likewise it may limit the local participant from fully understanding the intention of such message. To address such issue, the use of a humanoid telepresence robot, which faithfully acts as a surrogate, may be required. Such robot is ideal for use cases where high levels of immersion and tele-embodiment increase the success of a given task. An example of such use case can be seen in the use of a humanoid telepresence patrol robot [FIU - DISCOVERY LAB CITATION]. Due to the nature of the task, which involves high degrees of social interaction, the system must not only be mechanically and programmatically robust but also achieve a high perceptual level of embodiment on the local participants. The use of such humanoid telepresence robot that accurately replicates an operatorâĂŹs gestures and intentions would inherently require an advance control mechanism. Current approaches include the use of devices such as gesture capture sensors, integrated software hardware gesture sensing products such as the Microsoft Kinect and other hardware/mechanical interfaces such as joysticks. Looking into the future, the ideal control interface would perhaps be a brain-machine interface that detects a userâĂŹs intention of movements as well as the intention to disengagement through the readings of neurons firing - mind control. This would lead to perhaps the efficient and seamless control of avatars or mechanical systems just with thoughts [27,28,32,41]. EMG Interfaces Human-robot control interfaces have been proposed in the past decades, ranging from the use eye movements, speech and most commonly the use of EMG sensors. The use of each of these types of interfaces has their limitations. When biosignal or camera-based control interfaces are applied to embodied robotics or virtual avatars, the range of tracking areas may vary from a partial to total body [2,14,15,33,45]. Therefore, various movement patterns including specific hand and finger gestures may function as input control signals of tele-robotic operation. In such scenario, following body parts or whole body gesture would not be suitable to indicate engage or disengage commands. Furthermore, the use speech/voice is not always suitable for suitable noisy environments, and eye-based and even EMG based interfaces can be obtrusive. The application of EMG-based interfaces can be seen in literature that addresses the field of assistive technologies [30,48], where the user suffers from a medical condition that incapacitates them the use of other faculties such as their hands to control mechanical devices such as electric wheelchairs. Beyond assistive technologies, the literature addresses the use of EMG-sensors as the control interfaces for anthropomorphic robotic parts such as arms [3,21]. Furthermore, the use of tongue-based user interfaces has been explored [30,52], some of which rely on different technologies such optical sensors [43]. However, in this paper we solely focus on tongue gesture classification, using EMG sensors, with the intention of providing a mechanism for disengagement from a fully-immersive telepresence avatar. In comparison, to other EMG-based tongue classification systems which use up to 22 EMG-sensors for the classification of 6 tongue gestures [44,48], our system only uses only 5-EMG sensors to classify 6 tongue gestures with an accuracy of 94EMG Signal Classification Human Machine Interfaces (HMI) have extensively used EMG sensors to enable humans to control mechanical systems. These mechanical systems include, physical tools, robotic prostheses and even computer programs [36]. These electrical signals are formed by physiological variations in the state of muscle fibers. The combination of muscle fibersâĂŹ electro-potentials form what is known as a motor unit action potential (MUAP) [42]. EMG sensors non-invasively detect MUAP signals, which can later be used for diagnosis of illnesses or to analyze signal patterns. The analysis of such signal patterns enables the creation of systems that can identify muscle movement with a high degree of accuracy. To create such accurate systems the challenges posed by the field of signal processing must be addressed. These challenges include, but are not limited to, the removal signal noise influenced by the physiological and anatomical nature of muscles, as well as noise produced by the experimental setup [42]. During our experiment - due to the anatomy under the chin, our initial collection of data experienced the phenomenon known âĂIJcrosstalkâĂİ [13,51]. Essentially, unmonitored muscles contaminated the desired signal information. The latter was addressed by the relocation of the electrodes. It is worthy to note that other anatomical features affect the collection of signals, i.e. the amount of tissue between electrodes and contracting muscles or excess body fat [1,10]. After collecting the signal data, a preprocessing step often takes place which consists of segmenting the data, followed by filtering and rectification of the data. A series of high-pass and low-pass filters are available to address issues such as that of electromagnetics

noise, motion artifacts created by moving cables, and internal noise [36]. Following the preprocessing step, for each divided segment a feature set is computed by what is known as feature extraction. Feature extraction plays a crucial role in achieving better classification accuracy. This process involves the transformation of EMG signal into a feature vector that can be fed to a machine learning classifier. Various studies propose different feature extraction techniques in the time domain, frequency domain and time-frequency domain or a combination [50]. These techniques include mean absolute value (MAV), root mean square (RMS), autoregressive (AR), discrete and continuous wavelet transforms (DWT)(CWT) just to name a few [8,36,39]. Finally a machine learning algorithm is used to create a predictive model. Machine learning classifiers often used on EMG data include, support vector machines (SVM), feedforward artificial neural networks (ANN), recurrent neural networks (RNN), linear discriminant analysis (LDA) [8]. In effort to achieve better results, data reduction techniques such as principal component analysis (PCA) are often used.

METHODS

In this section, we describe ...

Data Collection

Your paper's title, authors and affiliations should run across the full width of the page in a single column 17.8 cm (7 in.) wide. The title should be in Helvetica or Arial 18-point bold. Authors' names should be in Times New Roman or Times Roman 12-point bold, and affiliations in 12-point regular.

See `\author` section of this template for instructions on how to format the authors. For more than three authors, you may have to place some address information in a footnote, or in a named section at the end of your paper. Names may optionally be placed in a single centered row instead of at the top of each column. Leave one 10-point line of white space below the last line of affiliations.

Data analysis

The acquired surface electromyography (sEMG) was preprocessed offline using MATLAB R2016a (The MathWorks Inc., Natick, Massachusetts, USA). The data (2 s per trial) were detrended to subtract mean values from the signals and filtered with a IIR comb notch filter to attenuate power noise and its harmonics. A band-pass filter (16th-order Butterworth, cutoff frequency between 10 and 500 Hz) was implemented to suppress high-frequency noises and artifacts due to movement. For feature extraction, a single preprocessed dataset from each channel was truncated into windows of 128 samples without overlapping based on time scaling, and the root-mean square (RMS) value, which denote the average sEMG amplitude, of each window was calculated according to the following equation:

Classification

Classification using SVM We used C-SVC (C-Support Vector Classification) algorithm implemented in LIBSVM [6]. LIBSVM implements the âĂIJone-against-oneâĂİ technique (Knerr et al., 1990) for multi-class classification; this technique reduces multi-class classification into multiple binary classification problem. The classification problem is described as follow: Let k be the number of classes, then (k(k-1))âĄĎ2 classifiers would constructed. Each classifier then trains data from two classes. Then, for training data from the i-th and j-th classes, we solved the following two-class classification problem.

subject to

C >0 is the penalty parameter of the error term. We use the radial basis function (RBF) described below: Radial basis function (RBF):

Here, Îş, is a kernel parameter. Notes on SVM As literature states, scaling is very important when using SVM. It is often recommended to linearly scale each attribute to the range [-1, +1] or [0, 1] [6]. Prior to building the model, the values obtained from the signal preprocessing step are normalized on the above scale. Note that this means that our test dataset had to be scaled prior to using it as input to test the model. For our use case, we chose the radial basis function as our kernel function. Due to the nonlinear correlation between attributes of our EMG data and the class label, the RBF kernel was chosen. The selection of our preprocessing step i.e. window selection affected our decision to choose the RBF kernel - since, the RBF kernel is not suitable for datasets with large number of features [6].

As noted by the function, the parameter that we need to "search" for ar: C and Îş. The goal then becomes to find such (C,Îş), which provide us with a classifier that predicts tongue gesture classes with high accuracy. Many computational methods exist for finding (C,Îş) [6,17]. For our use case, we chose to perform grid-search with a 10-fold cross validation. Overall, grid-search can be seen as a simplistic approach that entails doing an exhaustive parameter search through approximations [17]. As noted by Lin, C., since (C,Îş) are independent grid-search can be parallelized. Our dataset and source code can be found on [SOURCE CODE CITATION].

RESULTS

In this section, we ...

Table 2 shows the accuracy level of the classifier trained using each subjectâĂŹs data. The average accuracy of the tongue interface on both subjects is of 94%. Each individual classifiers provides an accuracy of 97Table 3 shows the confusion matrix of subject 1 which represents the classification accuracy of the tongue classifications. The highest misclassifications occur between on the forward tongue gesture, where the gesture is classified as an up or down gesture. This can be a result of the similarity between the tongue gestures and noise artifacts resulting from the execution of the gesture and placement of the EMG sensors.

DISCUSSION

Limitation

Closed-mouth In this experiment, tongue gestures were only classified with closed mouth. To be applied practically, the proposed EMG-based interface should be able to distinguish

between talking or unintentional tongue movements and a combination/sequence of intentional tongue motions with closed mouth for engaging/disengaging mode changes. Generalized model To further emphasize, the challenges of signal noise greatly affect the creation of a generalized classifier that can accurately classify tongue gesture across any subjects. For example, amplitude of sEMG can be decreased exponentially due to variant anatomical and biological factors such as increased distance between muscle fibre and electrodes, composition of fibre type, and muscle structure [1,9,10,31,34]. As mentioned in previous sections, moreover, EMG crosstalk often contaminate signals; thus, different placement of electrodes between subjects and even within same participant at different dates may introduce signals of different muscles [5,9]. Challenges related to signal noise must be carefully taken into consideration to develop a production-ready device. For instance, the design of device should be able to adjust electrode placements dependent on anatomical structures of the userâĂŹs under chin. Furthermore, the system also needs to be flexible with respect to different speed and path of movements between trials within same user or participants. This would then lead to a stable accuracy in prediction.

Future Work and Other Applications

Our future efforts are focused on (1) the creation of a generalized tongue-gesture control system, (2) identifying a user-friendly way to activate the system, (3) exploring a system, which uses an ensemble of different approaches, i.e. voice recognition along with an EMG-based approach. Further subsequent research is focused on merging such EMG interface with other control modalities, i.e. allowing an augmented reality overlay to be displayed after the system receives a tongue signal. Such display, for example, can then allow the operator to choose a range of options via the use of manual gestures or arm motions. Such options would include the modulation of display settings and volume while the operator is engaging in a virtual environment. In the near future, we plan to further develop the system, apply in virtual environment, and detect userâĂŹs intention in real-time. Combination of tongue gestures as an ultimate means of interface in our experiment will not only vary the number of commands but also create distinct patterns different from features derived by speaking or other voluntary tongue movements.

CONCLUSION

In conclusion, this paper presented an EMG control interface for the disengagement, of an immersed operator, from a remote or virtual environment. The interface consists of 5 EMG sensors, which are placed below the suprahyoid bone. More specifically the sensors are placed above the surface of the (1) geniohyoid and mylohyoid (above the hyoid bone), (2) left side of mylohyoid, (3) right side of mylohyoid, (4) between digastric (posterior belly) and stylohyoid on the left side, and (5) between digastric (posterior belly) and stylohyoid muscles on the right side. The readings from these sensors were then used to train a support vector machine classifier, which classified a userâĂŹs tongue gesture with an average 94% accuracy. A specific set of tongue gestures can then be translated to the userâĂŹs intention to disengage from the remote environment or virtual world.

ACKNOWLEDGMENTS

. Authors 1, 2, and 3 gratefully acknowledge the donation from Jeremy Robins for initiating the TeleBot project.

REFERENCES FORMAT

Your references should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible and may be obtained by any reader for a nominal fee. Proprietary information may not be cited. Private communications should be acknowledged in the main text, not referenced (e.g., [Golovchinsky, personal communication]). References must be the same font size as other body text. References should be in alphabetical order by last name of first author. Use a numbered list of references at the end of the article, ordered alphabetically by last name of first author, and referenced by numbers in brackets. For papers from conference proceedings, include the title of the paper and the name of the conference. Do not include the location of the conference or the exact date; do include the page numbers if available.

References should be in ACM citation format: http://www.acm.org/publications/submissions/latex_style. This includes citations to Internet resources [4, 3, 9] according to ACM format, although it is often appropriate to include URLs directly in the text, as above. Example reference formatting for individual journal articles [2], articles in conference proceedings [7], books [10], theses [11], book chapters [12], an entire journal issue [6], websites [1, 3], tweets [4], patents [5], games [8], and online videos [9] is given here. See the examples of citations at the end of this document and in the accompanying BibTeX document. This formatting is a edited version of the format automatically generated by the ACM Digital Library (http://dl.acm.org) as "ACM Ref." DOI and/or URL links are optional but encouraged as are full first names. Note that the Hyperlink style used throughout this document uses blue links; however, URLs in the references section may optionally appear in black.

REFERENCES

1. ACM. 1998. How to Classify Works Using ACM's Computing Classification System. (1998). http://www.acm.org/class/how_to_use.html.
2. R. E. Anderson. 1992. Social Impacts of Computing: Codes of Professional Ethics. *Social Science Computer Review December* 10, 4 (1992), 453–469. DOI: http://dx.doi.org/10.1177/089443939201000402
3. Anna Cavender, Shari Trewin, and Vicki Hanson. 2014. Accessible Writing Guide. (2014). http://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-writing-guide/.
4. @_CHINOSAUR. 2014. "VENUE IS TOO COLD" #BINGO #CHI2014. Tweet. (1 May 2014). Retrieved Febuary 2, 2015 from https://twitter.com/_CHINOSAUR/status/461864317415989248.

5. Morton L. Heilig. 1962. Sensorama Simulator. U.S. Patent 3,050,870. (28 August 1962). Filed Februrary 22, 1962.

6. Jofish Kaye and Paul Dourish. 2014. Special issue on science fiction and ubiquitous computing. *Personal and Ubiquitous Computing* 18, 4 (2014), 765–766. DOI: http://dx.doi.org/10.1007/s00779-014-0773-4

7. Scott R. Klemmer, Michael Thomsen, Ethan Phelps-Goodman, Robert Lee, and James A. Landay. 2002. Where Do Web Sites Come from?: Capturing and Interacting with Design History. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 1–8. DOI:http://dx.doi.org/10.1145/503376.503378

8. Nintendo R&D1 and Intelligent Systems. 1994. *Super Metroid*. Game [SNES]. (18 April 1994). Nintendo, Kyoto, Japan. Played August 2011.

9. Psy. 2012. Gangnam Style. Video. (15 July 2012). Retrieved August 22, 2014 from https://www.youtube.com/watch?v=9bZkp7q19f0.

10. Marilyn Schwartz. 1995. *Guidelines for Bias-Free Writing*. ERIC, Bloomington, IN, USA.

11. Ivan E. Sutherland. 1963. *Sketchpad, a Man-Machine Graphical Communication System*. Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA.

12. Langdon Winner. 1999. *The Social Shaping of Technology* (2nd ed.). Open University Press, UK, Chapter Do artifacts have politics?, 28–40.