Decoding Human Intent Using a Wearable System and Multi-Modal Sensor Data

Cemil S. Geyik, Arindam Dutta, Umit Y. Ogras and Daniel W. Bliss, *Fellow, IEEE*
School of Electrical, Computer, and Energy Engineering, Arizona State University
Email: {cemil.geyik, adutta7, umit, d.w.bliss}@asu.edu

***Abstract*—Despite the phenomenal advances in the computational power of electronic systems, human-machine interaction has been largely limited to simple control panels, such as keyboards and mice, which only use physical senses. Consequently, these systems either rely critically on close human guidance or operate almost independently. A richer experience can be achieved if cognitive inputs are used in addition to the physical senses. Towards this end, this paper introduces a simple wearable system that consists of a motion processing unit and brain-machine interface. We show that our system can successfully employ cognitive indicators to predict human activity.**

I. Introduction

The way in which humans interact with machines is rapidly evolving. The complexity of human involvement is sometimes hidden behind the seemingly simple interaction methods itself, which can use physical, cognitive or affective inputs [1]. The physical aspect is related to the mechanics of interaction. It is the most commonly used method today, since physical interaction allows for simple devices, such as keyboards, mice and displays. However, in this form of interaction, machines are passive agents that can only respond to physical inputs. Cognitive inputs can significantly improve the user experience by facilitating intellectual interaction, e.g., household electronics controlled by brain activity. Similarly, affective inputs can help machines understand emotional state of the user, e.g., a music player selecting the tracks based on user's mood. Consequently, we can enable machines to understand what their user wants by enriching current physical interaction with cognitive and affective inputs.

Pervasive use of cognitive and affective inputs for human-machine interaction relies critically on two technologies. First, there is a need for methodologies and algorithms that can process physiological signals, such as electroencephalogram (EEG), to decode the users' intentions and needs. To enable a seamless interaction, this processing has to be in situ and real-time. Therefore, the second requirement is a wearable system that can monitor and process the relevant physiological signals. This capability can enable a symbiotic human-machine relationship by providing a continuous interaction and real-time feedback between the human and machines.

This paper first presents a simple wearable system prototype capable of sensing, processing and communicating user motion. Together with a commercial brain machine interface (BMI) [2], this system enables us to analyze user gestures and brain activity reflected by EEG signals. In order to do so, we designed an experiment involving a repetitive task, i.e., moving small boxes from one side of a table to the other side. Before completing this task, the user may be given a second task to stimulate cognitive activity, e.g., drawing a picture. If the second task is given, the user starts performing it upon completing the repetitive task. Our objective is to determine whether the second task is given or not by monitoring cognitive activity. To achieve this objective, we developed a methodology to analyze the EEG signals with the help of timing information extracted from the hand movement data. We successfully show that EEG signals can be used as an indicator of future user activity. Therefore, this work serves as a first step towards a wearable system that can decode human intent using multi-modal sensor data.

The rest of the paper is organized as follows. Section II presents the related work. Section III presents the methodology for human intent decoding. Finally, Section IV discusses the experimental results, and Section V concludes the paper.

II. Related Research

The term human intent has been used to express the partial intent sufficient enough in the context of a specific target application. For example, in a car disassembly task, the intent is the desired direction and magnitude change in the position of the end-effector [3]. Similarly, a driver's turn intent [4, 5] is sufficient to represent the intent in driver assistance applications. In our context, human intent refers to whether the user is preparing to perform a second task upon completing the current one or not. It has been shown that the human intent can be predicted by low-level limb movements [6], vision processing [7], and using physiological signals [8]. There have been numerous attempts at human decoding intent by classification of single modalities, such as speech [9], body movement [10], head movement [11], gestures [12], facial expression [13], eye movement [14], hand pressure [15] and brain-activity [16]. However, single modality is not sufficient for robust human-machine communication [17], which is the first step towards symbiosis. Indeed, intuitive *human-human* communication leverages multiple modalities, such as speech, gestures, mimics, and body language [18]. In this work, we employ two modalities. The main modality is EEG processing with the help of the Emotiv EPOC+ EEG headset. To improve the effectiveness of the EEG signal processing, we also use a tri-axis gyroscope and accelerometer data which encodes the hand movements.

Brain machine interfaces (BMI) have become a promising technology to interface directly with computers [19]. Impressive progress in novel BMI technologies is discussed in [20]. The first symbiotic BMI was introduced by Mahmoudi and Sanchez [21]. This approach has an adaptive actor interface to the motor cortex and a critic that evaluates the value of actor's actions using Nucleus Accumbens. The authors show that a continuous perception-action-reward cycle enables operation in changing environments by adapting the BMI decoder.

III. Multi-Modal Intent Decoding Methodology

A. Overview

The proposed multi-modal intent decoding methodology is illustrated in Figure 1. The flow consists of hand motion processing and EEG signal processing. Hand motion is monitored using a wearable prototype described in Section III-B. The purpose of hand motion processing is to provide timing information about the user activity. More precisely, we segment the EEG data non-uniformly into multiple blocks with the help of hand motion processing. EEG processing itself consists of conditioning the raw data, segmentation and feature extraction using spatiotemporal eigenspectrum construction. The following subsections detail each step of this flow.

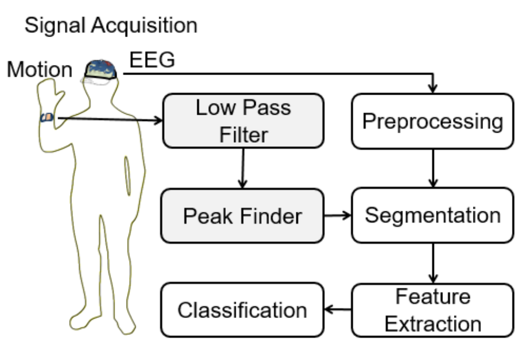
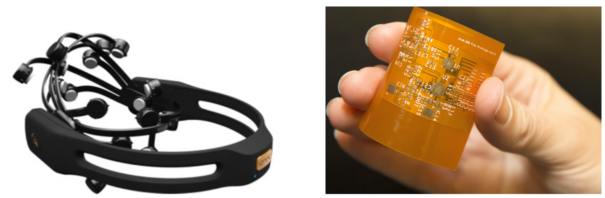


Fig. 1: Multi-modal intent decoding methodology flow.

B. Wearable System Prototype

We record the brain activity using Emotiv EPOC+ EEG headset [2] shown in Figure 2. The Emotiv headset uses 14 channels and two reference electrodes The first reference is an absolute voltage reference called common mode sense (CMS). The second one, referred to as driven right leg (DRL), is a feedback cancellation system to float the reference level on the common mode body potential. The raw EEG data is stored using wireless connection at a sampling rate of $128Hz$.

The user hand motion was recorded at a sampling rate of $100Hz$ using a motion processing unit that integrates a tri-axis gyroscope and accelerometer, and a TI-CC2650 microcontroller [22] on a flexible polyimide substrate. It senses motion, processes the raw data locally, and supports both Bluetooth low energy and Zigbee wireless communication protocols. The maximum supported data rate is $192kbps$, and the average power consumption is $12.2mW$ at $1.44kpbs$



(a) (b)

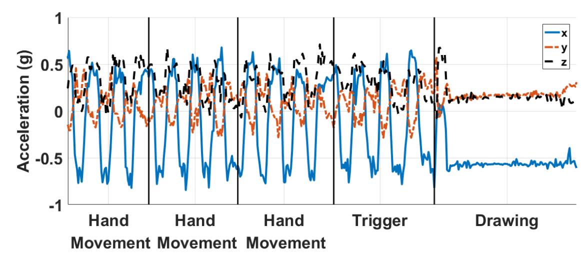
Fig. 2: (a) Emotiv EPOC+ headset, (b) motion processing unit.

throughput. Our prototype device occupies $3.8cm \times 3.8cm$ and weighs $50mg$ including the programming interfaces. Its area and low weight allow us to attach it directly to the skin or clothing. Detailed design specification can be found in [23].

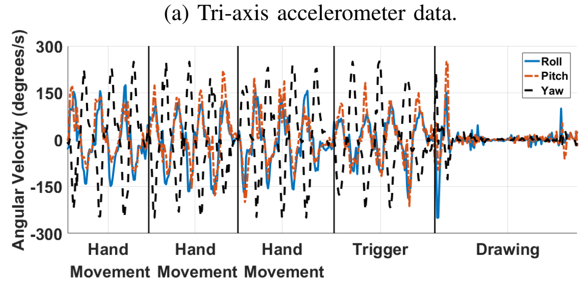
C. Hand Motion Processing

As a reminder, the repetitive task in our experiment is moving a set of boxes on a table from one side to the other side. The repetitive task is said to be complete, if all the boxes are moved from one side to the other. To infer the start time of each new round, we employed accelerator and gyroscope measurements. The accelerator gives acceleration along $x-$, $y-$ and $z-$ axis in terms of gravitational constant g. Gyroscope measures the rate of rotation, i.e., angular velocity, in three components called roll, pitch and yaw. The raw tri-axial accelerometer and gyroscope readings are plotted in Figure 3(a) and Figure 3(b), respectively.

Both gyroscope and accelerometer readings are preprocessed using a 5-point moving average filter to suppress the measurement noise. Then, the filtered data is used to find the angular velocity of the hand. The local minima of the angular velocity indicate the start time of each new round. We find the local



(a) Tri-axis accelerometer data.



(b) Roll, pitch and yaw components of the gyroscope data.

Fig. 3: Raw motion data that encodes the hand movements.

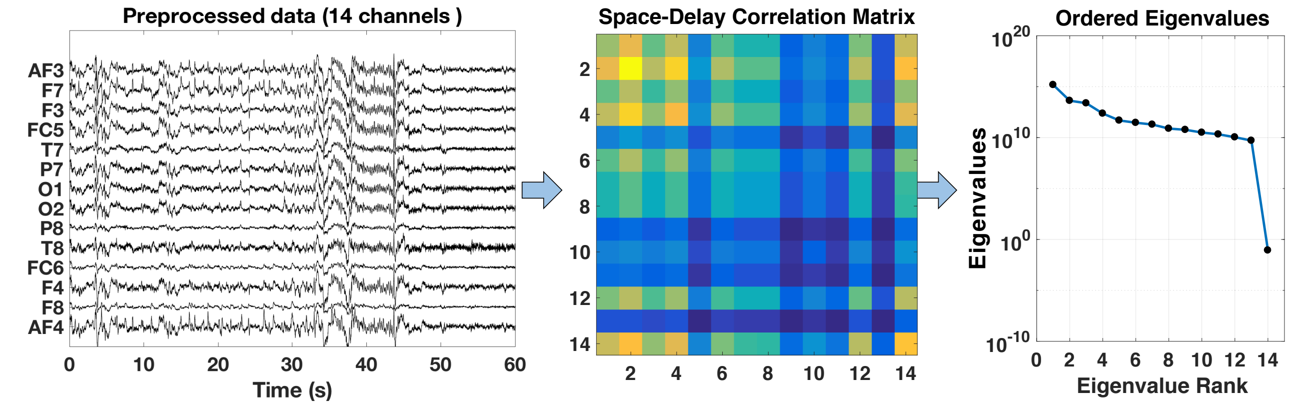


Fig. 4: Flow diagram for eigenspectrum computation. The EEG signals along the y-axis of the left-most graph belong to 14 different channels. More information about the electrode locations corresponding to each labels can be found at [2].

minima using a peak finder algorithm, and identify the start times, as illustrated by the vertical lines in Figure 3.

D. EEG Signal Processing

To remove the irrelevant artifacts from the measure data, we first filter the raw EEG data using a high pass filter with a pass band starting at $4Hz$ and a notch filter at $60Hz$. Then, we use the time instances from motion processing to segment the input data into multiple blocks. Let T_s be the sampling interval and n_c be the number of EEG channels. The EEG signal sampled at time kT_s in block b can be represented as:

$$z_b(k) = [z_{b,1}(k),\ z_{b,2}(k), \ldots, z_{b,n_c}(k)] \quad 0 \leq k < n_s \qquad (1)$$

where each column represents a different channel and n_s is the number of samples. Hence, the measured signal is represented by a $n_s \times n_c$ matrix, i.e., $z_b(k) \in R^{n_s \times n_c}$.

Spatiotemporal correlation has been proven to be an effective method to extract features from EEG data [24]. Therefore, we use $z_b(k)$ first to construct the space-delay data matrix. Then, we use the resulting matrix to compute the space-delay correlation matrix. Let us denote the time-delayed multichannel signal as $z_b(k - d\tau)$, where d is the delay scale and τ is the delay amount. By using multiple delay scales, we can construct the space-delay data matrix for block b as:

$$\mathbf{X}_{bd} = [z_b(k - 1\tau),\ z_b(k - 2\tau), \ldots, z_b(k - d\tau)] \qquad (2)$$

Using d time delay scales grows the dimension of the space-delay matrix to $n_s \times dn_c$. This helps characterizing the spatiotemporal correlation efficiently over a long span of relative time [24]. Let $g(\cdot)$ be a function that shifts and normalizes each column of its input to zero mean, unit variance. We approximate the spatiotemporal correlation matrix for block b $\mathbf{R}_b$ using the normalized zero-mean data as:

$$\mathbf{R}_b = \frac{1}{n_s} g(\mathbf{X}_{bd})^T g(\mathbf{X}_{bd}) \qquad (3)$$

Spatiotemporal Eigenspectrum – Next, we compute the eigenvalues of the spatiotemporal correlation matrix $\mathbf{R}_b$. Rank order of the eigenvalues gives the spatiotemporal eigenspectrum, which encodes invaluable information about the cognitive activity as a function of time (i.e., blocks) as illustrated by our experiments. The whole process starting from the EEG signals to eigenspectrum computation is summarized in Figure 4 for reference.

Parameters Used in this Work – In this work, we used 14 EEG channels ($n_c = 14$) sampled at $T_s = \frac{1}{128}s$ intervals. We used three different delay scales ($d = 3$), and minimum delay amount of $\tau = 1$ sample. Finally, the EEG data is segmented nonuniformly into 5 blocks using the data obtained from motion processing. The first three blocks correspond to the repetitive task performed by the user, the fourth block contains the trigger event, and the last block covers the cognitive task.

IV. Experimental Results

A. Experiment Setup

During our experiments, subjects wore the Emotiv EPOC+ headset and the motion processing prototype. Then, the subject remained seated on a chair in front of a desk, as shown in Figure 5. We placed four small boxes on the one side of the table, and asked the subject to move all boxes from one side to another. The task is considered complete when all the boxes are moved to the opposite side. The subject repeated the whole task until s/he was asked to stop, or draw an arbitrary object on a piece of paper upon completing the current task. Four neurologically healthy subjects participated in the study.

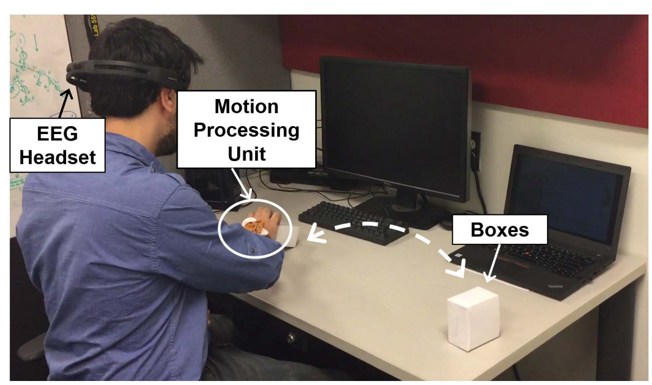


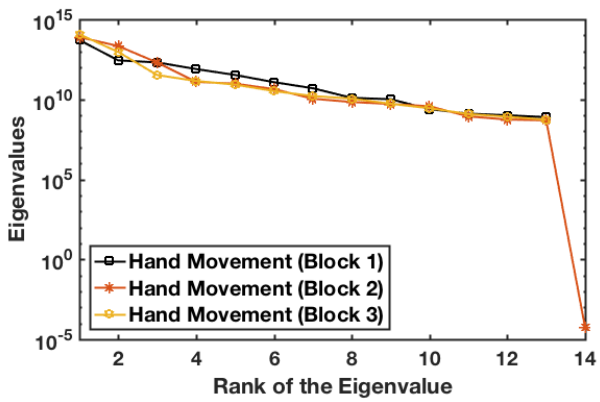
Fig. 5: The experimental setup.

Subjects were not under the influence of any medication that could interfere with EEG. The execution of the repetitive task and drawing lasted approximately $50s$, and each subject repeated the experiment five times.

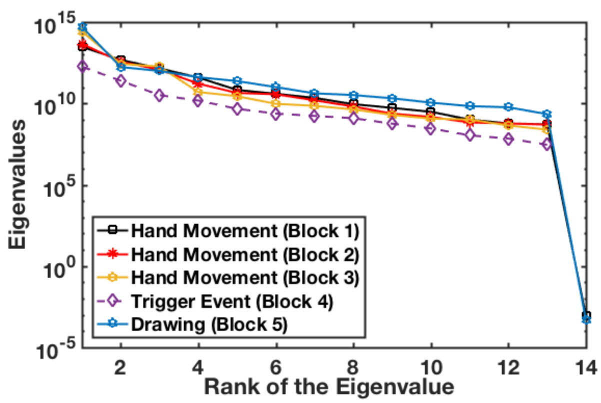
B. Analysis of the EEG Data

We first used the accelerometer and gyroscope data to identify the start and end times of each repetitive task. Then, this information is used to divide the EEG data into multiple blocks labeled as *hand movement*, *trigger event* and *drawing*. Hand movement corresponds to the repetitive task. If the subject is asked to draw an object, we refer to the corresponding block as the trigger event. Finally, if the subject draws an object, the corresponding block is labeled as drawing.

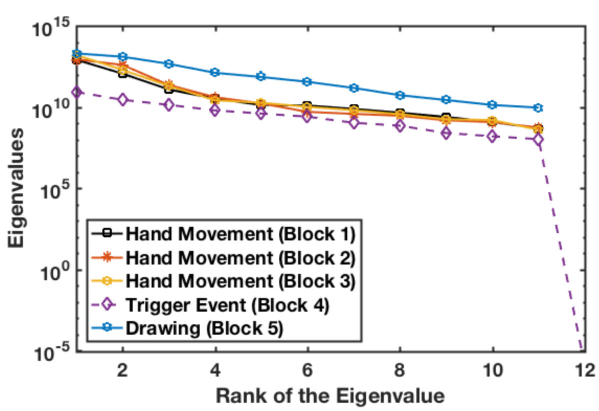
The eigenvalues of the space-delay correlation matrix are shown in Figure 6. In particular, Figure 6(a) plots the eigenvalues for Subject 1, when he performed only the repetitive task. We observe that the eigenvalues follow a very similar pattern for each of block. Eigenvalues for two other subjects, who were asked to draw an object, are plotted in Figure 6(b) and Figure 6(c). The eigenvalues show a very clear distinction, when the trigger happens. This means that the subject starts thinking about the drawing before completing the repetitive task. This is clearly reflected in the eigenvalue distribution of the fourth block. Similarly, the spectra shown in Figure 7 shows that the trigger event has a distinctive signature. Each column in these plots correspond to a different block of time, while each row shows the eigenvalues at a given rank. All three blocks in Figure 7(a) have a similar behavior as they all correspond to repetitive task as in Figure 6(a). However, the spectra that correspond to the trigger event can be easily distinguished from the other blocks, as shown in Figure 7(b) and Figure 7(c). The



(a) Subject 1: Performed *only* the repetitive hand movement.

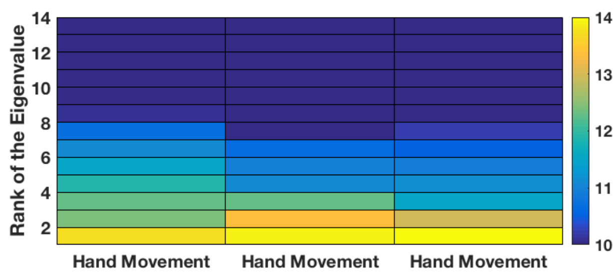


(b) Subject 2: Performed both the repetitive and drawing tasks.

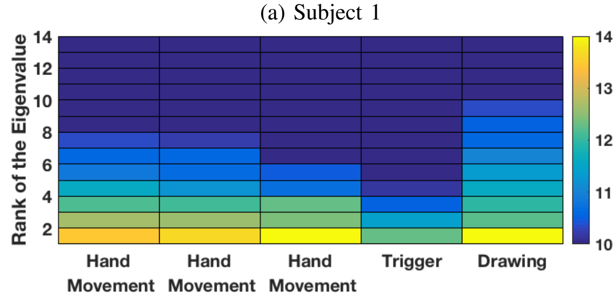


(c) Subject 3: Performed both the repetitive and drawing tasks. The EEG channels were removed in this experiment due to sensor misplacement.

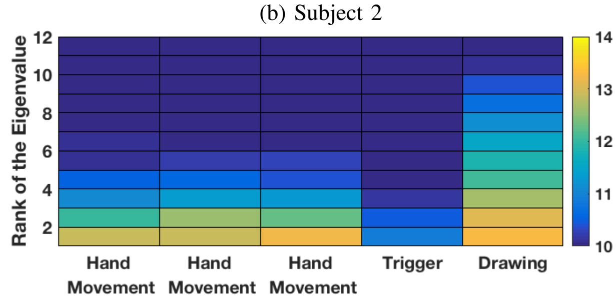
Fig. 6: Space-delay eigenspectra for three different subjects.



(a) Subject 1



(b) Subject 2



(c) Subject 3

Fig. 7: Eigenspectrogram of three different subjects.

last spectrum, that corresponds to Subject 3, used 12 out of 14 EEG channels, since two channels were discarded due to sensor displacement. This shows that the proposed approach can also work effectively with fewer channels.

V. Conclusions and Future Work

This paper presented a methodology for cognitive activity detection to predict future human action for human intent decoding. The proposed approach employs a wearable system that can monitor hand movements and sense EEG signals. Hand movement is used to divide the EEG signals into multiple blocks. For each block, the eigenspecta extracted from space-delay covariance matrix is used as to classify multi-channel EEG data. Our experiments show that the cognitive activity can be used successfully to predict future human activity.

Acknowledgment

The authors would like to thank Rohit Voleti and Owen Ma of School of Electrical, Computer and Energy Engineering, Arizona State University, for their assistance and suggestions that greatly improved this paper.

References

[1] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1):137–159, 2008.

[2] Emotiv. EEG Neuroheadset. https://www.emotiv.com/support/ (accessed December 3, 2016).

[3] Mustafa Suphi Erden and Tetsuo Tomiyama. Human-Intent Detection and Physically Interactive Control of a Robot without Force Sensors. *Robotics*, 26(2):370–382, 2010.

[4] Shinko Yuanhsien Cheng and Mohan Manubhai Trivedi. Turn-Intent Analysis using Body Pose for Intelligent Driver Assistance. *Pervasive comp.*, 5(4):28–37, 2006.

[5] Anup Doshi and Mohan Trivedi. A Comparative Exploration of Eye Gaze and Head Motion Cues for Lane Change Intent Prediction. In *Intelligent Vehicles Symposium*, pages 49–54, 2008.

[6] Chrystopher L Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, and Rachid Alami. A Methodological Approach Relating the Classification of Gesture to Identification of Human Intent in the Context of Human-robot Interaction. In *Robot and Human Interactive Comm.*, pages 371–377, 2005.

[7] Erik Murphy-Chutorian and Mohan Manubhai Trivedi. Head Pose Estimation in Computer Vision: A Survey. *Pattern Analysis and Machine Intelligence*, 31(4):607–626, 2009.

[8] Dana Kulic and Elizabeth A Croft. Affective State Estimation for Human–Robot Interaction. *Robotics*, 23(5):991–1000, 2007.

[9] Heriberto Cuayáhuitl. Robot Learning from Verbal Interaction: A Brief Survey. *New Frontiers in Human-Robot Interaction*, page 62, 2015.

[10] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective Body Expression Perception and Recognition: A Survey. *Affective comp.*, 4(1):15–33, 2013.

[11] AMER Al-Rahayfeh and MIAD Faezipour. Eye Tracking and Head Movement Detection: A State-of-Art Survey. *Translational Eng. in Health and Med.*, 1:2100212–2100212, 2013.

[12] Siddharth S Rautaray and Anupam Agrawal. Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.

[13] Gabriel Castaneda Oscos, Taghi M Khoshgoftaar, and Randall Wald. Rotation Invariant Face Recognition Survey. In *Info. Reuse and Integration (IRI)*, pages 835–840, 2014.

[14] Andreas Bulling. Human Visual Behaviour for Collaborative Human-Machine Interaction. In *Proc. of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous comp. and Proc. of the 2015 ACM Int. Symposium on Wearable comp.*, pages 901–905, 2015.

[15] Ross David McLachlan. *Pressure as a Non-dominant Hand Input Modality for Bimanual Interaction Techniques on Touchscreen Tablets*. PhD thesis, University of Glasgow, 2015.

[16] Luzheng Bi, Xin-An Fan, and Yili Liu. EEG-based Brain-Controlled Mobile Robots: A Survey. *Human-Machine Sys.*, 43(2):161–176, 2013.

[17] Md Muztoba, Ujjwal Gupta, Mustofa Tanvir, and Umit Y Ogras. Robust Communication with IoT Devices using Wearable Brain Machine Interfaces. In *Proc. of Int. Conf. on Computer-Aided Design*, Nov. 2015.

[18] Rejeev Sharma, Vladimir I Pavlovic, and Thomas S Huang. Toward Multimodal Human-computer Interface. *Proc. of the IEEE*, 86(5):853–869, 1998.

[19] Lin Li, Austin J Brockmeier, John S Choi, Joseph T Francis, Justin C Sanchez, and José C Príncipe. A Tensor-Product-Kernel Framework for Multiscale Neural Activity Decoding and Control. *Computational Intelligence and Neuroscience*, 2014:2, 2014.

[20] Robbin A Miranda, William D Casebeer, Amy M Hein, Jack W Judy, Eric P Krotkov, Tracy L Laabs, Justin E Manzo, Kent G Pankratz, Gill A Pratt, Justin C Sanchez, et al. DARPA-Funded Efforts in the Development of Novel Brain-Computer Interface Technologies. *Journal of Neuroscience Methods*, 244:52–67, 2015.

[21] Babak Mahmoudi and Justin C Sanchez. A Symbiotic Brain-Machine Interface through Value-based Decision Making. *PloS one*, 6(3):e14760, 2011.

[22] Texas Instruments Inc. CC-2650 Microcontroller. http://www.ti.com/product/CC2650.

[23] Ujjwal Gupta, Jaehyun Park, Hitesh Joshi, and Umit Y. Ogras. Flexibility Aware Systems on Polymer: Concept to Prototype. *IEEE Trans. on Multi-Scale Computing Systems*, 2017. (*To appear.*).

[24] James R Williamson, Daniel W Bliss, David W Browne, and Jaishree T Narayanan. Seizure prediction using eeg spatiotemporal correlation structure. *Epilepsy & Behavior*, 25(2):230–238, 2012.