Prediction Error Negativity in Physical Human-Robot Collaboration

Authors*

***Abstract*— Cognitive conflict is a fundamental phenomenon of human cognition, particularly during interaction with the real world. Understanding and detecting cognitive conflict can help to improve interactions in a variety of applications such as in human-robot collaboration (HRC), which involves constantly guiding the semi-autonomous robot to perform a task in given settings. There have been several works to detect cognitive conflict in HRC but without physical control settings. In this work, we have conducted the first study to explore cognitive conflict using prediction error negativity (PEN) in physical human-robot collaboration (pHRC). Our results show that there is statistically significant (p = .047) higher PEN for conflict condition compared to normal condition, as well as a statistically significant difference between different levels of PEN (p = .020). These results indicate that cognitive conflict can be detected in pHRC settings and, consequently, provide a window of opportunities to improve the interaction in pHRC.** *(Abstract)*

Keywords- EEG; pHRC; Cognitive Conflict; PEN

I. Introduction

Cognition can be defined as a process of acquiring information through thought, experience, and senses. Intellectual functions such as reasoning, memorising, thinking, and learning is considered as part of cognition [1]. However, there is another aspect of intellectual function that has been studied extensively by researchers, that is, the ability to predict a result using experiences from the past. Whenever humans are performing a task, the human brain is constantly predicting the actions and outcomes of that task [2]. This prediction process that simulates the response of a task to estimate the outcome is known as the internal model [3]. However, inevitably, there will be a mismatch between the internal model and reality, and this phenomenon is known as cognitive conflict [4]. Studies about cognitive conflict have been investigated for a long time — the first research conducted about three decades ago by Falkenstein et al. [5] in 1991, found a negative response in event-related potentials (ERP) [6] using electroencephalogram (EEG) in a bimanual-choice reaction task, this response is known as error-related negativity (ERN). After this first work, there have been several works that came in light using different terminology depending on the task that is performed. For example, visual mismatch negativity (vMMN) is used to define cognitive conflict in the experiment of visual effect monitoring [7], and N2 or N200 is another ERP feature elicited by visual stimuli [8]. Feedback related negativity (FRN) had demonstrated its presence in the experiment of price or reward prediction. Recently, prediction error negativity (PEN) was discovered in an experiment involving active motor movement [9]. Even if PEN is part of the same family of negativity, due to being caused by an erroneous behaviour, it is significantly different in terms of behaviour and is invoked during the task rather than after the end of the task or its observation.

Understanding and detecting cognitive conflict can help to improve interactions in a variety of applications. One such application is human-robot collaboration (HRC), which involves guiding a semi-autonomous robot to perform a task in given settings. In any HRC, there are several possible instances where a user might invoke cognitive conflict due to the unexpected behaviour of the robot. There have been several experiments conducted to evaluate cognitive conflict during HRC. The researchers conducted studies where participants needed to observe and judge the actions performed by a robot as either correct or incorrect [10, 11]. The authors found negativity around 400 ms in ERP after incorrect action. They also developed a reinforcement learning method in a way that minimizes cognitive conflict information. In another study [12], an experiment using the EEG signal conducted to use ERN to trigger the safety measure of a robot when a human operator observed an unexpected event during a robot navigation task. In a study [13], the participant needed to observe robotic actions, which included random context-dependent erroneous behavior. The authors detected ErrP (Error-related potentials) in ERP caused by erroneous behavior of a robot. Similar work shows in [14], where authors utilize the ErrP due to the erroneous task to correct the robot behaviour. Although these works have great potential, participants observe and/or control the robot passively (limited motor movement), which is not common in real-world settings. In the passive state experiment, the participants either only observe robot movement from a distance or provide input to the robot using button pressed on keyboards/joysticks [15]. In our proposed work, however, the participants performed a task that requires active physical human-robot collaboration (pHRC), i.e. active physical control of the semi-autonomous or guided robot to perform a task. Studies that address interactions based on active motor movement are rare in EEG related works due to several other possible confounding factors and artifacts [16]. A study by Singh et al. [9] successfully provided the first proof to find cognitive conflicts in a scenario that involved active motor movement. They conducted an experiment where the participants performed a 3D object selection task with tracked hand motions in an immersive virtual reality (VR) environment and found a new signature of cognitive conflict in ERP known as PEN. In the line of this research, we have conducted a pilot experiment in pHRC settings, and our results show [17] clue toward the existence of similar negativity effect in pHRC.

Following the pilot experiment, we have conducted further experiments to investigate the cognitive conflict caused by forces exchanged during pHRC tasks. We hypothesized that erroneous behaviour of a robot during pHRC will invoke PEN in ERP.

II. Experiment and Methodology

A. Participants

EEG data were recorded from ten participants (including two females). The mean age of the participants was 30, with a range of 25-39 years, and no history of any psychological disorders. All participants were instructed not to consume any drinks or drugs containing chemicals that can alter brain activity at least three hours before the experiment. Following the explanation of experimental procedures, all participants provided their consent before participating in the experiment. This experiment obtained approval of this human research ethics committee at the University of Technology Sydney (ETH18-3029).

B. Experiment Setup

In this EEG based experiment, a 32-channel wireless EEG system named MOVE (Brain Product GmbH, Germany) has been used. The placement of the EEG electrodes was consistent with the 10-20 international system [18]. The contact impedance is maintained below 25kΩ, and EEG recordings were digitally sampled at 1000 Hz.

The collaborative robotic system, ANBOT[18], was used to perform the task. It features a UR10 robot arm by Universal Robots [19], and, between the robot end-effector and the robotic arm, an ATI mini-45 transducer is mounted to measure and control the user interaction forces and torques. The participant is in contact with the robot arm through a handle placed on the robot end-effector. Figure 1 shows how a participant collaborates with ANBOT to perform the experiment.

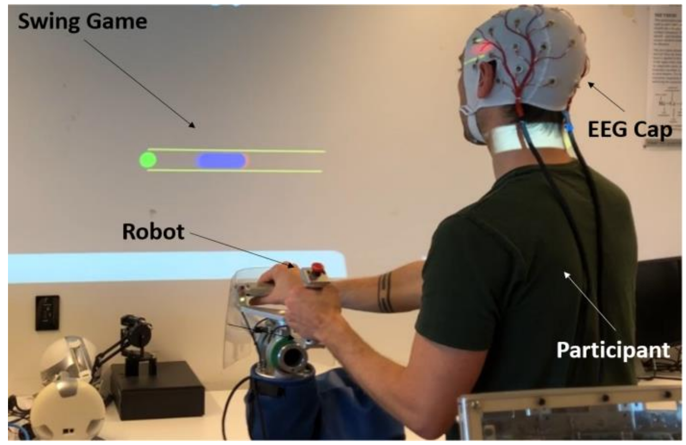

Figure 1. A participant performing a pHRC task while wearing an EEG cap

C. Experiment design

The experiment task is presented as an interactive game called the 'Swing Game'. It is a mono-directional game that was designed to be as simple as possible in order to avoid requiring too much input from participants and allow them to focus on the task. There are three possible targets located at the center and both sides of the bar. The green circle represents the current target, where the participant has to move the blue circle, while the red circles represent the potential target. To keep the participant more engaged and use both hands on the task, a paint sprayer scenario was emulated. The blue circle represents the point where the nozzle of the robot is aiming. The participant needed to hold the press button located at the back handle of the robot end-effector in order to move the robot arm, and the blue circle while holding the press button located at the front handle will paint the area blue where the blue circle is. In addition, a scoring system that calculates the score depending on the sprayed area and the distance from the target when the trial ends is implemented to keep the participants engaged in the game.

Each trial always starts from the center, and one of the two targets will turn from red to green randomly. Participants have five seconds to complete the trial, or else the trial will reset. The center target will turn green to inform the participant to move the blue circle back, after which the next trial is started. An invisible obstacle will be placed halfway to the target with a probability of 40% and stops the motion of the robot completely. Since the obstacle is invisible, participants will experience cognitive conflict once they hit the obstacle during the experiment.

As per Figure 2, it can be seen that there are four different conditions in this game; these were sudden force normal, sudden force obstacle, smooth force normal and smooth force obstacle. Sudden normal and sudden obstacle conditions presented a resistive force right before the target or the invisible obstacle, whereas in the smooth normal and smooth obstacle conditions, the resistance started earlier. The sudden force obstacle and smooth force obstacle represent conflict conditions while sudden force normal and smooth force normal represent normal (non-conflict condition). For each condition, there was one session of 125 trials, therefore for four-session 500 trials. Participants spent about 120-150 minutes for the whole experiment, including EEG preparation time.

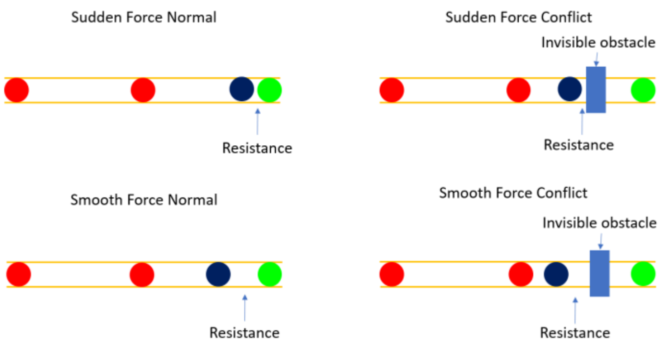

Figure 2. Experiment design: The participant moves the blue circle (user location) from red circle (start location) in the center to randomly left or right to green circle (ending location) as a task in four conditions of sudden force normal, sudden force conflict, smooth force normal, and smooth force conflict.

D. Data processing

EEG data processing was performed using EEGLAB [19], a toolbox in MATLAB (Mathwork Inc, USA). Raw EEG data were filtered using a 1Hz high-pass and a 50Hz low-pass finite impulse response (FIR) filter. Another set of data is also filtered using a 2Hz high-pass and a 50Hz low-pass FIR filter. After filtering, the data were down-sampled to 250Hz and subjected to the visual inspection of the noisy channels. After the removal of some obvious noisy channels, an EEGLAB function called automatic channel rejection [20] is used to remove remaining noisy channels. Subsequently, the data is re-referenced to average. Artifact Subspace Reconstruction (ASR) [21, 22] is then applied to the 2-50hz dataset, followed by Independent Component Analysis (ICA) [23]. After the completion of ICA, ICA information matrices (icawinv, icasphere, and icaweights) from the 2-50hz dataset is then copied to the 1-50hz dataset in order to avoid low-level frequency noise. Independent components (ICs) and other noise are then rejected using a function in EEGLAB called ADJUST [24]. After the rejection of the bad component, dipole fitting [25] is applied to the dataset in order to obtain the residual variance. Any component with a residual variance greater than 15% is also removed. Epochs are then extracted from 500ms before the onset of the resistive force (either before the obstacle or before the target, i.e. green circle) to 1000ms after the same force. After epochs were extracted, data smoothing, detrending, and a 4-8hz bandpass FIR filter are applied to epochs in order to remove data linear trend from the data and filter within the theta range [26]. Based on previous work [9], the ERP of each epoch is then extracted at electrode location Fz, which is a midline frontal EEG channel. The PEN was calculated by the average of the minimum peak happening in the time range between 150ms-250ms in the extracted ERP from epochs for all conditions. A flowchart of the preprocessing pipeline is shown in Figure 3 below.

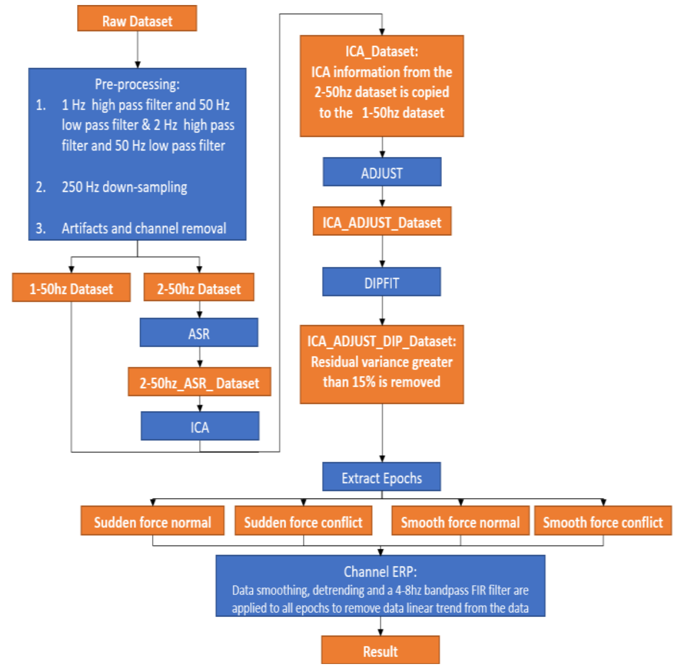

Figure 3. Data processing pipeline

All the statistical analysis has been performed using SPSS (IBM Inc.) over-extracted PEN from EEG data analysis.

III. Results and Discussion

We have evaluated the ERP based on 'Fz', which was found to be the prominent EEG channel selection in previous studies [9, 17]. It can be seen in Figure 4 that a sudden obstacle and smooth obstacle conditions clearly show negative deflection (PEN) around 150-250 ms followed by a positive increase (Pe) around 250-400ms. For the sudden normal and smooth normal conditions PEN and Pe are not very obvious and visible. The negative deflection in ERP i.e. PEN occurs when participants could not move blue circle toward the green circle due to an invisible obstacle in sudden and smooth conditions. But in the no obstacle condition (sudden normal and smooth normal), the participants were able to move the blue circle from the center to the green circle without any obstacle, therefore no visible PEN appears in ERP. Although all four conditions show clear Pe, which is in line with other research related to visual feedback [27, 28].

To further understand the negative deflection in all four conditions, we have extracted the PEN amplitude from 150-250ms for all individuals from their averaged ERP. It can be clearly seen from Table 1 that 70% of participants show higher PEN amplitude for conflict conditions (sudden and smooth obstacle) compared to normal conditions (sudden and smooth normal). We also analyzed if there is any difference between sudden obstacle and smooth obstacle conditions. There was a higher PEN for sudden obstacle condition compared to smooth obstacle condition.

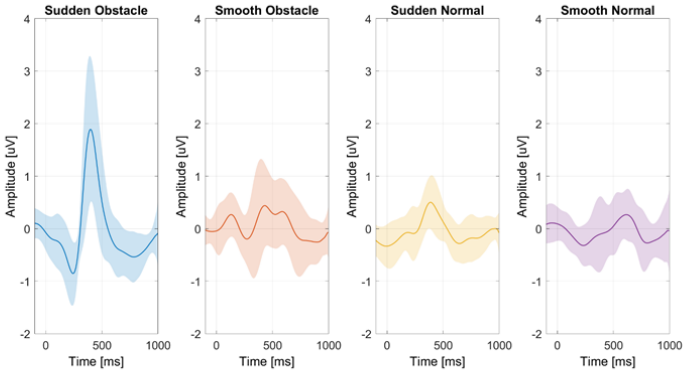

Figure 4. Average ERP for all conditions of sudden obstacle, smooth obstacle, sudden normal, and smooth normal

TABLE I. PEN AMPLITUDE FOR ALL CONDITIONS AND ALL PARTICIPANTS (SU-OB : SUDDEN OBSTACLE; SM-OB : SMOOTH OBSTACLE; SU-NOR : SUDDEN NORMAL; SM-NOR: SMOOTH NORMAL)

Participants	**Amplitude (μV)**			
	Su-Ob	**Sm-Ob**	**Su-Nor**	**Sm-Nor**
S1	-1.0786	-0.0066	-0.2839	-0.6513
S2	-0.4405	0.0200	-0.05954	-0.04802
S3	-0.2999	-0.2190	-0.2994	-0.2097
S4	-1.4791	-0.3895	-0.4473	-0.6795
S5	-1.7164	-1.0034	-1.1080	-0.4853
S6	-0.6397	-0.2417	0.3086	-0.6679
S7	-1.1466	-0.2015	-0.2420	-0.4035
S8	-1.6187	-0.4788	-0.2107	0.3536
S9	-0.3459	0.3043	-0.0308	-0.9644
S10	-1.8357	-0.6131	-0.2148	-0.0763

To understand if variation in PEN among all conditions holds statistically, we have performed repeated-measures ANOVA on 2 X 2 experiment design conditions, such that two types of force (sudden and smooth) were combined with two types of conditions (normal and obstacle). The statistical results show that there was statistically significant ($F(1,9) = 5.280$, $p = .047$) main effect for sudden and smooth conditions and also statistically significant difference ($F(1,9) = 8.000$, $p = .020$) for normal and obstacle conditions. We also looked at the interaction between force (sudden and smooth) and conditions type (obstacle and normal), and it was found that the interaction between force and conditions type is also statistically significant ($F(1,9) = 24.160$, $p= .001$).

To further understand whether such statistical difference exists, we have performed post hoc comparison using the paired t-test. The test revealed that there was a statistically significant difference ($p = .001$) between sudden obstacle and sudden normal conditions, as well as statistically significant difference ($p < .001$) between sudden obstacle and smooth obstacle. Although there was no statistically significant difference ($p = .510$) between smooth obstacle and smooth normal condition.

The results from this experiment support our hypothesis that cognitive conflict can be measured by PEN in pHRC settings, where humans and robots actively exchange forces, and not just in HRC. Even if PEN shows similar behaviour as other categories of negativities [29-31], the cognitive process involved in pHRC requires continuous information from visual and proprioceptive feedback [32, 33] to perform the task. In pHRC, the participants do not observe and decide if the action was correct or wrong at the end of the task, but instantaneously adapt their strategy, even before finishing the task. This happens also in the presence of the invisible obstacle. Those results open up many questions related to cognitive processes involved in pHRC and similar settings.

IV. CONCLUSION

We have conducted a study to find if the cognitive conflict can be detected in pHRC. The Swing Game has been designed with sudden and smooth forces combined with normal and obstacle conditions. We have found a significant difference in PEN in normal and obstacle conditions for both sudden and smooth forces. We also found that PEN is significantly higher for sudden force conditions compared to smooth force conditions. These results conclude that cognitive conflict represented by PEN can be detected in pHRC settings and potentially can be used to improve the interaction and collaboration.

ACKNOWLEDGMENT

This study is supported in part by the Australian Research Council (ARC) Linkage Project [*] and the ARC Discovery Project Grant [*].

REFERENCES

[1] Z. W. Pylyshyn, *What is cognitive science?*

[2] M. S. Ryoo, "Human activity prediction: Early recognition of ongoing activities from streaming videos," in *2011 International Conference on Computer Vision*, 2011: IEEE, pp. 1036-1043.

[3] B. A. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457-465, 1976.

[4] J. Cooper, *Cognitive dissonance: 50 years of a classic theory*. Sage, 2007.

[5] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke, "Effects of crossmodal divided attention on late ERP components. II. Error processing in

choice reaction tasks," (in eng), *Electroencephalography and clinical neurophysiology*, vol. 78, no. 6, pp. 447-55, Jun 1991.

[6] P. L. Nunez and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

[7] C. Stagg, P. Hindley, A. Tales, and S. Butler, "Visual mismatch negativity: the detection of stimulus change," *NeuroReport*, vol. 15, no. 4, pp. 659-663, 2004. [Online]. Available: https://journals.lww.com/neuroreport/Fulltext/2004/03220/Visual_mismatch_negativity__the_detection_of.17.aspx.

[8] B. Kopp, F. Rist, and U. W. E. Mattler, "N200 in the flanker task as a neurobehavioral tool for investigating executive control," *Psychophysiology*, vol. 33, no. 3, pp. 282-294, 1996/05/01 1996, doi: 10.1111/j.1469-8986.1996.tb00425.x.

[9] A. K. Singh *et al.*, "Visual Appearance Modulates Prediction Error in Virtual Reality," *IEEE Access*, vol. 6, pp. 24617-24624, 2018, doi: 10.1109/ACCESS.2018.2832089.

[10] I. Iturrate, L. Montesano, and J. Minguez, "Single trial recognition of error-related potentials during observation of robot operation," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010: IEEE, pp. 4181-4184.

[11] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using EEG-based reward signals," in *2010 IEEE International Conference on Robotics and Automation*, 2010: IEEE, pp. 4822-4829.

[12] C. I. Penaloza, Y. Mae, M. Kojima, and T. Arai, "Brain signal-based safety measure activation for robotic systems," *Advanced Robotics*, vol. 29, no. 19, pp. 1234-1242, 2015.

[13] S. Ehrlich and G. Cheng, "A neuro-based method for detecting context-dependent erroneous robot action," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016: IEEE, pp. 477-482.

[14] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017: IEEE, pp. 6570-6577.

[15] M. Spüler and C. Niethammer, "Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity," (in English), *Frontiers in Human Neuroscience*, Original Research vol. 9, no. 155, 2015-March-26 2015, doi: 10.3389/fnhum.2015.00155.

[16] A. Puce and M. S. Hämäläinen, "A Review of Issues Related to Data Acquisition and Analysis in EEG/MEG Studies," *Brain sciences*, vol. 7, no. 6, p. 58, 2017, doi: 10.3390/brainsci7060058.

[17] S. Aldini *et al.*, "Effect of Mechanical Resistance on Cognitive Conflict in Physical Human-Robot Collaboration," in *International Conference on Robotics and Automation*, 2019.

[18] M. G. Carmichael; *et al.*, "The ANBOT: An Intelligent Robotic Co-worker for Industrial Abrasive Blasting," in *2019 IEEE International Conference on Intelligent Robots and Systems (IROS)*, Macao, 2019: IEEE.

[19] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9-21, 3/15/ 2004, doi: http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

[20] G. Gómez-Herrero, "Automatic artifact removal (AAR) toolbox v1. 3 (Release 09.12. 2007) for MATLAB," *Tampere University of Technology*, 2007.

[21] C. A. E. Kothe and T. P. Jung, "Artifact removal techniques with signal reconstruction," ed: Google Patents, 2016.

[22] C.-T. Lin, C.-S. Huang, W.-Y. Yang, A. K. Singh, C.-H. Chuang, and Y.-K. Wang, "Real-Time EEG Signal Enhancement Using Canonical Correlation Analysis and Gaussian Mixture Clustering," *Journal of Healthcare Engineering*, vol. 2018, p. 11, 2018, Art no. 5081258, doi: 10.1155/2018/5081258.

[23] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in neural information processing systems*, 1996, pp. 145-151.

[24] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229-240, 2011, doi: doi:10.1111/j.1469-8986.2010.01061.x.

[25] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data," *Computational Intelligence and Neuroscience*, vol. 2011, p. 9, 2011, Art no. 156869, doi: 10.1155/2011/156869.

[26] P. Luu, D. M. Tucker, and S. Makeig, "Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation," *Clinical Neurophysiology*, vol. 115, no. 8, pp. 1821-1835, 2004/08/01/ 2004, doi: https://doi.org/10.1016/j.clinph.2004.03.031.

[27] S. H. Patel and P. N. Azzam, "Characterization of N200 and P300: Selected Studies of the Event-

Related Potential," *International Journal of Medical Sciences,* vol. 2, no. 4, pp. 147-154, 10/01
08/08/received
09/15/accepted 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1252727/.

[28] H. M. Gray, N. Ambady, W. T. Lowenthal, and P. Deldin, "P300 as an index of attention to self-relevant stimuli," *Journal of Experimental Social Psychology,* vol. 40, no. 2, pp. 216-224, 3// 2004, doi: http://dx.doi.org/10.1016/S0022-1031(03)00092-1.

[29] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin, "The Error-Related Negativity," *Perspectives on Psychological Science,* vol. 13, no. 2, pp. 200-204, 2018/03/01 2018, doi: 10.1177/1745691617715310.

[30] G. Stefanics, P. Astikainen, and I. Czigler, "Visual mismatch negativity (vMMN): a prediction error signal in the visual modality," (in English), *Frontiers in Human Neuroscience,* Editorial vol. 8, no. 1074, 2015-January-22 2015, doi: 10.3389/fnhum.2014.01074.

[31] E. S. Kappenman, S. J. Luck, W. J. Gehring, Y. Liu, J. M. Orr, and J. Carp, "The Error-Related Negativity (ERN/Ne)," ed.

[32] B. Lee *et al.*, "Engineering Artificial Somatosensation Through Cortical Stimulation in Humans," (in English), *Frontiers in systems neuroscience,* Original Research vol. 12, no. 24, 2018-June-04 2018, doi: 10.3389/fnsys.2018.00024.

[33] R. J. van Beers, A. C. Sittig, and J. J. Gon, "Integration of proprioceptive and visual position-information: An experimentally supported model," (in eng), *Journal of neurophysiology,* vol. 81, no. 3, pp. 1355-64, Mar 1999.