sciendo

Journal of Smart Internet of Things (JSIoT)
VOL 2023, No.02 | 105-114 | 2023
DOI: 10.2478/jsiot-2023-0013

Article

Gaze Tracking for Hands-Free Human Using Deep Reinforcement Learning Approach

Irfan Ullah[1], Abid Ali[1]*, Shahid Rasool[1], Abdul Moiz Khan[1], Iqra Batool[1], Manahil Javed[1], and Sarara Kalsoom[1]

[1] Electrical Engineering Department, Gomal University, D.I.Khan, Pakistan

*Corresponding author: Abid Ali (email: artificiali19@gmail.com)

Received 23 August 2021; Accepted 17 September 2021; Published 15 December 2023

Abstract: People with disabilities have new and advanced methods to communicate with the applications for virtual keyboards and other communication tools. In this paper, we utilized a novel deep reinforcement learning-based technique for determining the location of the accessible options for gaze-controlled tree-based menu selection system. A virtual English keyboard has been incorporated into the layout of the new user interface, which also includes improved requests for text modification through the gaze. The two methods used to manage the system are: 1) eye tracking for typing on the virtual keyboard and 2) eye tracking with a device for soft-switch utilizing deep reinforcement learning. Simulation results show that DRL based algorithm outperforms other baseline techniques in terms of total loss and accuracy.

Keywords: Gaze Tracking, Deep reinforcement learning, Eye gaze, Hands-Free.

Copyright © 2023 Journal of Smart Internet of Things (JSIoT) published by Future Science for Digital Publishing and Sciendo . This is an open access article license CC BY (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

More than 20 million people experience a stroke each year, and because they are unable to move or speak, they often struggle to communicate successfully [1]. Innovative, and effective communication strategies based on augmentative with alternative ways have been created to assist these people [2]. In the systems mentioned above, [3], a wide range of physical, alphabetical, and symbolic signals are frequently incorporated. Patients with aphasia and tetraplegia, among others, have serious speech and mobility problems, and this field of study is accelerating to meet their demands [4]. The two primary components of alternative and augmentative systems are communication software and input device(s). For example, input devices for electromyography and/or electrocardiograph signals are used to record any voluntary intent from patients. One of the preferred input devices is the eye-tracker, which captures the user's eye gaze in real-time [5]. Eye-gaze controlling systems have already been developed for people with locked-in syndrome [6] and neurodegenerative illnesses [7]. In addition to eye-gaze and non-invasive electroencephalography

(EEG) recordings, brain-computer interfaces have also been utilized to communicate [8], [9]. Additionally, it has been proposed to use specialized communication tools, such as onscreen keyboards [10], [11] and complex communication spreadsheets [12], to analyze the data acquired by input devices and produce explicit information about the users' intent.

The keyboard layout of a common virtual keyboard is based on a specific language [13]. This type of keyboard, designated [14] is among the simplest and most fundamental tools for text entering alternatively. The key design cue for an efficient virtual keyboard is the corresponding language script [15]. The majority of virtual keyboards currently available were built with the English language in mind, and optimization algorithms were developed in accordance with this syntax [4], [16].

In this paper, we introduce a novel reinforcement learning method [17] and a deep neural network (DRL) for tracking the gaze of stroke patients or persons using a hands-free device. DRL has many advantages over sensor-based serving since it replaces a designed control mechanism with a trial-and-error learning model. A feedback signal showing whether human (considered as agent) behaviors are advantageous and may be used while optimizing the reward-based function using a deep neural network or not. For instance, the agent may gradually improve its behavior. The model may be taught both offline and online and has exceptional adaptive capabilities. As will be discussed in more detail below, there is no requirement for an annotated training dataset, which is typically the case with other machine learning techniques.

We adapt the gaze tracking problem into the setting of RL and propose a reward function based on the temporal order of cameras and audio observations that is currently available. The LSTM (a neural network recurrent model)-based deep networks we present allow us to examine both early and late integration of auditory and visual information. The action-value function is described by Deep RL. We provide a virtual model that eliminates the need for arduous interaction that takes hours to complete and enables us to train the specified deep RL model. By conducting trials on a publicly available dataset and on a real system, we provide empirical confirmation that our technique achieves cutting-edge performance.

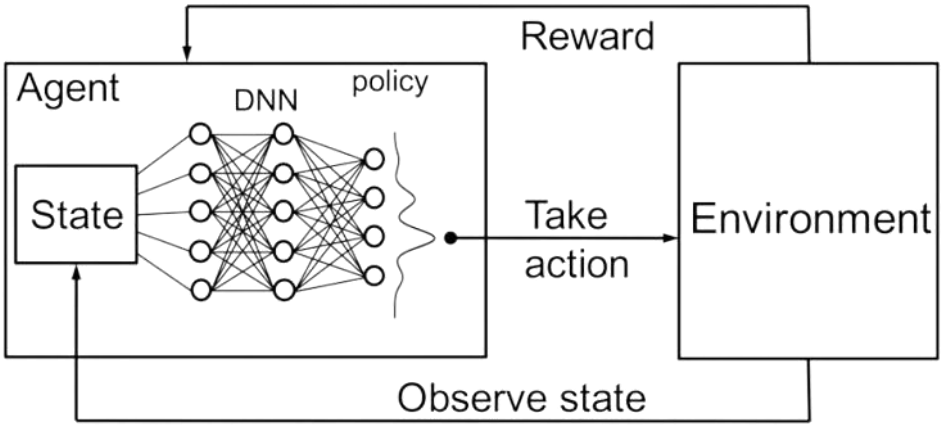


Figure 1. A Basic DRL Approach.

1.1 Related Works

Recent studies showed that text entry rates utilizing the Dasher keyboard were significantly faster than those using the dwell keyboard [18], [19]. Dasher's input rate was nearly twice as fast as Eye-Typing's. The initial training input rates for Dasher and Eye-Typing were 6.6 wpm and 4.6 wpm, respectively. Entry rates increased with experience. Since the last session, the input rates for Dasher and Eye-Typing have increased to 12.4 wpm and 6.0 wpm, respectively.

An eye-tracking-only and eye-tracking-in-conjunction-with-soft-switch multimodal virtual keyboard system for English typing, presented in [20], respectively observed typing speeds of 9.3 and 15.26 lpm (alphabets per minute). A helmet-mounted gaze-controlled text input interface has been developed for the other language, with a typing speed of 9.63 wpm [21]. However, due to the hardware's incompatibility with individuals who are disabled, the system's usability has been questioned. Additionally, a word prediction method was used to evaluate the system with a tiny sample size (only 6 healthy volunteers), all while disregarding typing errors. Words per minute can be increased by using the word prediction technique (wpm). However, studies have shown that word-finishing and word-prediction techniques affect typing speed [22], [23].

Eye-tracker-based virtual keyboard interfaces encounter considerable challenges while simultaneously pointing and picking items using gaze and dwell time durations. On the other hand, selecting a low dwell time length may increase typing speed, but it may induce erratic eye movements, which could lead to inaccurate item selections. On the other hand, if the dwell time is excessively long, it can worsen the person's condition. This issue is known as the Magic Touch problem [24] in detection theory. Furthermore, the majority of early works focused primarily on developing original concepts for graphical user interfaces. However, their optimization has mostly been ignored when used with various methods of operating an interference device, such as eye-gaze and soft-switches.

In the past, a frequency-based letter arrangement was employed in the design of a virtual keyboard to enhance the user experience for the English language. This system was tested using a soft-switch gadget with four separate scan times (1200 ms, and 1500 ms, 1800 ms, 2100 ms,). The performance did not, however, include a typing speed specification. For the frequency-based organizational system and the alphabetical layout, only task execution time was recorded in [25]. This particular study demonstrates how a frequency-based organization strategy can reduce the time required to complete a task. Soft-switch and eye-tracking input modalities were also tested with this upgraded virtual keyboard design. Using eye-tracking with dwell period (1.5s) and eye-tracking with a soft-switch, respectively, it was discovered that the average typing speed was 12 lpm and 7.25 lpm [26].

The idea of a sensor-based automaton serving gaze control was previously discussed. For instance, developing a loop that aligns an object's observed position with its desired placement is required for visual serving [27]. This implies that people are aware of the straight and inverted Jacobian devices. To estimate these Jacobians instead, one may utilize reinforcement learning [28]. Recent advances in sensor-based serving have made it possible to control the audio modality directly linked to measured acoustic parameters. However, this technique relies heavily on the existence of a single continuous sound source, such as [29] [30]. Unfortunately, multi-person conversations do not fall under this. There are currently no sensor-based serving techniques that can connect with numerous individuals, integrate visual and audio components, and do so. When there are different modalities and, consequently, various types of sensors, it could be difficult to integrate the available sensory data effectively and to build an optimal controller based on hand-written rules that must take into consideration all potential circumstances.

Several human interaction strategies based on reinforcement learning are appropriate to our job. In [31], a reinforcement learning method is used to train a system to play a game with a partner. The computer program chooses the motor commands based on input from the eyes and force/torque feedback. The uncertainty associated with human behavior is simulated by a Gaussian process, and Bayesian optimization is utilized to select the best course of action at each time step. In [32], reinforcement learning is applied to change motion speed, timing, interaction distances, and gaze that are connected to human interaction. How much the subject moves and how long they stare during a particular engagement affect the reward. Since external cameras are required, this cannot be easily deployed in scenarios where humans must continue learning in the real world. Additionally, only one human subject may be used for the treatment. In [33], reinforcement learning is used to improve human interaction, and a system is trained using a human-provided incentive. This idea of interactive

reinforcement learning is advanced in [34] using a table-cleaning system. Visual and speech detection are used to ask advice from a trainer who appears to be a parent so that the system can quickly develop a solid policy.

The authors in [35] instruct students to focus a camera on the conversation's active speaker using an extrinsic reward. The system can readily detect a card that the engaged speaker raises, and it uses visual cues to present the prize. To determine where to point the camera, audio data is used. Authors in [36] and [37] both advocate the use of a multimodal deep Q-network to create interactions that are comparable to those of humans. To get in touch with a person, the device must select what to do. In all other situations, if the system tries to send any signal but is unsuccessful, the reward must either be negative, positive, or null. The system features a sensor that activates the reward when in use, and acquiring this skill takes fourteen training days. The closest work to ours is the system discussed in [38], where a reinforcement learning technique learns great rules to control the orientation of a mobile system during social group interactions. Eventually, the system picks up on the speaker's direction and turns its head in that direction. Their hypothetical data, however, is restricted to a few planned occurrences, static people, and a predetermined geographical distribution of the group, and it is used to create their model.

1.2 Contributions and Organization

. The main contributions of this paper is summarized as follows:

- The location of the tree-based menu items for gaze-controlled systems is proposed using a novel deep reinforcement learning-based technique.
- Deep reinforcement learning is used for eye tracking for typing on the virtual keyboard and eye tracking with a device for soft-switch to manage the system.

The following sections are included in the remainder of this paper. Section II provides the novel deep reinforcement learning-based tracking or control of gaze. In addition, Section III uses simulations to assess how well the proposed work performs. Section IV contains the analysis and conclusion.

2. Novel DRL-Based Tracking and Control of Gaze

We imagine a patient who ought to be looking at a crowd but is instead not utilizing their hands. As a result, the patient will have to develop a gaze control method on their own through trial and error. Eye movement (tracking) and user interface interaction are the intended patient behaviors. The patient should also pick an option from a multi-level menu that allows them to select any item in real-time while using a glaze. There will be ambiguity in the use of the word agent. It's crucial to keep in mind that our environment is suitable for using a Markov Decision Process. Markov Decision Processes are environments where nearly all RL challenges may be represented. Their main shortcoming is a lack of prior exposure to environmental dynamics.

The primary focus of reinforcement learning is the environment-based learning process. The observations of an agent with specific capabilities who can function in the environment have an impact on this type of learning. According on the activities and particular condition, the agent is rewarded. After the environment, states, and potential actions have been identified, a proper selection of incentives can help direct learning to achieve a desired objective. The goal of the algorithm or agent is to maximize the total reward. This could result in a loss of short-term profits. The whole DRL process is made up of few matrices that are given below:

Environment: The environment is the real setting in which an agent operates. Patients can engage or respond in this environment. In our example, the environment largely consists of an eye tracker for a virtual keyboard and a soft touchscreen for deliberate hand execution. To make the work of recognizing this tracking more manageable, we incorporate the set of (x, y) locations looked at earlier during interpretation, as recorded by eye tracking, into our environment. This group of (x, y) points inside that image make up the gaze plot. The entire universe in this instance is made up of the 2D virtual keyboard

and the overlay gaze selection. A thing that operates in the environment and reaps rewards is called an agent. In our case, the agent is a patient glaze, and our ultimate goal is for the patient glaze to be within reach of the virtual keyboard. Each patient serves as a DRL agent in the proposed DRL technique, where RL is based on the Markov decision process. Additionally, states are given with S, actions with A, and rewards with R [S; A; R].

State: The state, on the other hand, conveys the agent's current position. The status in this case relates to the location of the patient glaze, which is currently within the virtual keyboard's viewing area. In order to narrow the state space or various directions the point may go while looking for the keyboard to choose an option from the menu, we restrict our agent to only move along the gaze plot, or among pixels that the patient glanced at. The keyboard was most likely noticed at least once during the preceding simulated image interpretation, it is assumed. The glaze movement that successfully accesses a virtual keyboard is Sa, and the hand motion used to make a selection is Sb.

$$S = [S_a, S_b] \tag{1}$$

Action: Then, we define that action as a state transition. In this context, "action" refers to eye or glazing movement, particularly within certain regions of the gaze graph. The Policy, on the other hand, specifies what to do in each case. The purpose of DRL training is to develop a perfect policy. The patient (agent) in our scenario should move as quickly as possible near the option in a menu and stay there to mark or forecast that option. Additionally, the value is the whole future advantage that an agent will receive from carrying out a particular action in a particular circumstance. Moving away from the low value and toward the selection from a menu has a high value in our system. The likelihood that the keypad tying and menu selection will be chosen makes up the DRL agent's action space A.

Reward: We build our incentive system to align the agent's goal of maximizing cumulative reward in order to meet our goal of having the agent type on a virtual keyboard utilizing Glaze and select an item from a list. The proposed DRL-SL approach is presented in Figure. 2.

Offline DNN training and online reinforcement learning are the two phases of DRL. The DNN is trained offline to approximate the correlation between the state-action pair (S, A) and function Q(S, A), which reflects the expected reward with the discount of choosing to remain in-state S and taking action A. Given as follows is the function Q(S, A):

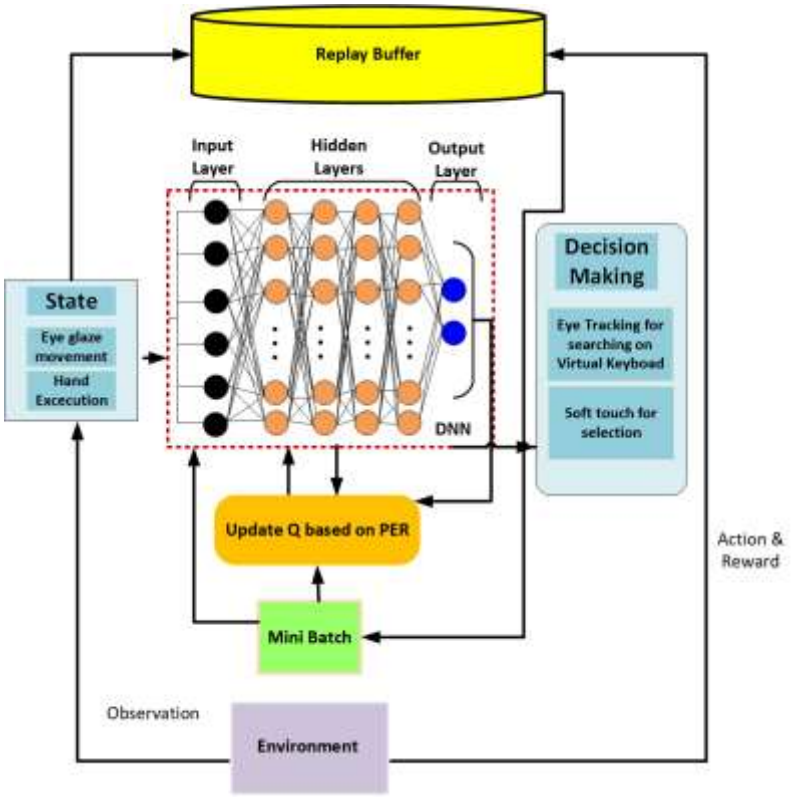


Figure. 2 A proposed DRL Approach.

$$Q(S,A) = E[r + \mu \max_{a'} Q(S', A') \mid S, A]; \tag{2}$$

Where the r and μ represents the reward and discount, respectively. The offline DNN training phase must gather sufficient experience samples at each epoch i where $E_i = (S_i, A_i, R_i, S_{i+1})$ presents is the experiences. Adding new experiences to the DNN that are randomly selected from the pool of recorded data and consist of state, action, instant reward, and state transition. In this offline method, the use of experience storage can smooth learning and avoid oscillations and parameter divergence. The control of the online probability built on the offline-made DNN is then improved using deep Q-learning. In each decision epoch t_i, the DRL agent retrieves the estimated Q from the DNN using the data of the current state S_i and each possible action A_i. Choosing the course of action with the highest Q value is our policy of greed. The DRL process then chooses the set R associated with the highest output value of the neural network, that is,

$$R = \arg\max_{R} Q(S_t \mid R; \theta) \tag{3}$$

Where θ stands for the weight of the deep neural network and $Q(S_t|R;)$ is the approximation of $Q(S_t, R)$ at time t, often known as a discounted average reward. The formula for the discounted average reward is

$$\begin{aligned} &Q(S_t, R) = E\{\sum_{n=t}^{\infty} \mu^{n-t} r_t \mid S_t, R\} \\ &\rightarrow Q(S_t, R) + \alpha[r + \mu \arg\max_{R} Q(S_{t+1}, R) - Q(S_t \mid R)] \end{aligned} \tag{4}$$

Where the range of the discount factor is shown as $0 \le \mu \le 1$. The deep neural network's weights are adjusted using the Bellman equation to reduce the mean squared error that results. As a result, the deep neural network can produce $Q_e(S_t|R;)$, which is an estimate of $Q(S_t|R)$. Using stochastic gradient descent, the loss function was defined as

$$L = E\{(r + \mu \arg\max_R Q_e(S_{t+1} \mid R;\theta) - Q_e(S_t \mid R;\theta))^2\} \quad (5)$$

The weights in the deep neural network can be changed in accordance with the results of the aforementioned mean square error (loss function) minimization.

$$\theta \leftarrow \theta \leftarrow \eta \delta_\theta L \quad (6)$$

Where δ_θ the generalized delta rule for gradient descent learning is applied and η is the learning rate.

3. Simulation Results

To our knowledge, DRL has not been directly applied to gaze-controlled based computer interaction, despite being widely employed in many cutting-edge applications, like robotics and gaming. This paper shows the first proof-of-concept use of DRL/RL on controlling gaze. By doing this, we can get a preview of the DRL's enormous potential for gaze controlling. We think that this strategy will ultimately serve to be a fundamental advance in machine learning, one that will usher in the era of actual natural language processing.

The method was capable of predicting gaze location on a different testing set with 85% accuracy after training using DRL on only few training photos. When learned on thousands or hundreds of training data set images, frequently with data augmentation, this illustrates the typical range of performance. The network here immediately started to overfit the data for training after being trained using the same training set. In the end, it had an accuracy of less than 10% in predicting the site of the testing set lesion. Given the widely recognized necessity for vast volumes of labeled data to perform relatively well, this last conclusion is not surprising.

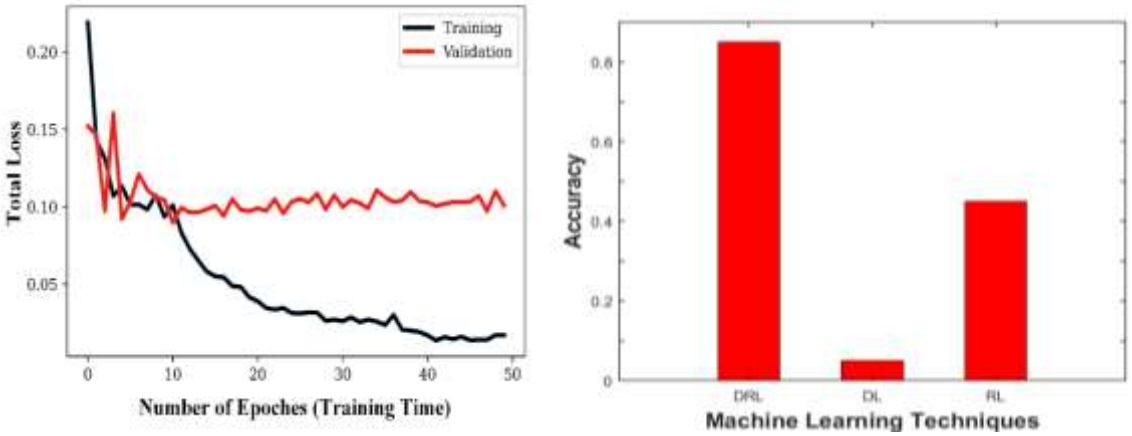
For two primary reasons, we are certain that DRL will revolutionize AI in gaze controlling.

1) Although when trained on extremely small data sets, it can produce correct findings.

2) Offer user perception about the photos being researched and profit from it.

Regarding the first advantage, we point out that 85% accuracy for a training batch of 70 photos is absolutely impressive. It is noteworthy that we did not even apply data augmentation, a common strategy for extending the size of the training sample. This capability offers two additional significant benefits in addition to the apparent advantages of expediting artificial intelligence research by making it possible for small data to power efficient artificial intelligence.

In Figure 3(a), for purposes of comparison, we employed a key-point recognition convolutional neural network with an overall similar structure to the network used in our DQN to train a supervised deep learning based on the convolutional neural network for 50 iterations. The validation dataset has already diverged well before 10th epoch, as seen in the image, and the system is starting to overfit the data for training more and more. In Figure 3(b), we performed the comparison of the predictions made on the testing set by a trained deep reinforcement learning (DRL), reinforcement learning (RL) and deep learning (DL). The proposed novel DRL based algorithm shows around 85% accuracy, whereas the RL and DL shows accuracy around 45% and 5%, respectively.



(**a**) (**b**)

Figure 3. Performance matrix to compare proposed DRL-SL with MD and OR: (**a**) Total Loss Vs Number of Epochs (Training Time); (**b**) Accuracy Vs Machine Learning Techniques.

4. CONCLUSIONS

Applications for virtual keyboards and other communication aids have given people with impairments new ways to communicate. In this work, we propose a novel deep reinforcement learning-based method for locating the displayed items for gaze-controlled tree-based menu selection systems. The structure of the new user interface has been built for an English virtual keyboard based on a menu in addition to better demands for text editing through staring. There are two ways to operate the system: eye tracking with a virtual keyboard and a deep reinforcement learning-based device for a soft switch.

References

[1] World Health Organization. (2011). Global Report. Accessed: Sep. 9, 2016. [Online]. Available: http://www.who.int/disabilities/worldreport/2011/en/.
[2] Wilkinson, K. M., & Hennig, S. (2007). The state of research and practice in augmentative and alternative communication for children with developmental/intellectual disabilities. Developmental Disabilities Research Reviews, 13(1), 58–69.
[3] DeCoste, D. C., & Glennen, S. (1997). The Handbook Augmentative Alternative Communication. San Diego, CA, USA: Singular.
[4] de Sousa Gomide, R., Loja, L. F. B., Lemos, R. P., Flˆores, E. L., Melo, F. R., & Teixeira, R. A. G. (2016). A new concept of assistive virtual keyboards based on a systematic review of text entry optimization techniques. Research in Biomedical Engineering, 32(2), 176–198.
[5] Mele, M. L., & Federici, S. (2012). A psychotechnological review on eye-tracking systems: Towards user experience. Disability and Rehabilitation: Assistive Technology, 7(4), 261–281.
[6] Park, S.-W., Yim, Y.-L., Yi, S.-H., Kim, H.-Y., & Jung, S.-M. (2012). Augmentative and alternative communication training using eye blink switch for locked-in syndrome patient. Annals of Rehabilitation Medicine, 36(2), 268–272.
[7] Cipresso, P., et al. (2011). The combined use of brain-computer interface and eye-tracking technology for cognitive assessment in amyotrophic lateral sclerosis. In Proceedings of the IEEE International Conference on PervasiveHealth (pp. 320–324).
[8] Schalk, G., Brunner, P., Gerhardt, L. A., Bischof, H., & Wolpaw, J. R. (2008). Brain–computer interfaces (BCIs): Detection instead of classification. Journal of Neuroscience Methods, 167(1), 51–62.
[9] Usakli, A. B., & Gurkan, S. (2010). Design of a novel efficient human-computer interface: An electrooculagram based virtual keyboard. IEEE Transactions on Instrumentation and Measurement, 59(8), 2099–2108.
[10] Fu, Y.-F., & Ho, C.-S. (2009). A fast text-based communication system for handicapped aphasiacs. In

Proceedings of the IEEE Conference on Information Intelligence and Security (pp. 583–594).
[11]Orhan, U., Hild, K. E., Erdogmus, D., Roark, B., Oken, B., & Fried Oken, M. (2013). RSVP keyboard: An EEG based typing interface. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1–11).
[12]Biswas, P., & Samanta, D. (2008). Friend: A communication aid for persons with disabilities. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16(2), 205–209.
[13]Molina, A. J., Rivera, O., & G´omez, I. (2009). Measuring performance of virtual keyboards based on cyclic scanning. In Proceedings of the IEEE International Conference on Autonomic and Autonomous Systems (pp. 174–178).
[14]Ghosh, S., Sarcar, S., & Samanta, D. (2011). Designing an efficient virtual keyboard for text composition in Bengali. In Proceedings of the ACM International Conference on Human-Computer Interaction in India (pp. 84–87).
[15]Bhattacharya, S., & Laha, S. (2013). Bengali text input interface design for mobile devices. Universal Access in the Information Society, 12(4), 441–451.
[16]Samanta, D., Sarcar, S., & Ghosh, S. (2013). An approach to design virtual keyboards for text composition in Indian languages. International Journal of Human-Computer Interaction, 29(8), 516–540.
[17]Sutton, R. S., & Barto, A. G. (1998). Introduction to Reinforcement Learning. 1st ed. MIT Press.
[18]Ward, D. J., & MacKay, D. J. C. (2002). Artificial intelligence: Fast hands-free writing by gaze direction. Nature, 418, 838.
[19]Rough, D., Vertanen, K., & Kristensson, P. O. (2014). An evaluation of dasher with a high-performance language model as a gaze communication method. In Proceedings of the International Working Conference on Advanced Visual Interfaces (pp. 169–176).
[20]Cecotti, H. (2016). A multimodal gaze-controlled virtual keyboard. IEEE Transactions on Human-Machine Systems, 46(4), 601–606.
[21]Sarcar, S., & Panwar, P. (2013). Eyeboard++: An enhanced eye gaze-based text entry system in Hindi. In Proceedings of the ACM International Conference on Computer-Human Interaction (pp. 354–363).
[22]Anson, D., et al. (2006). The effects of word completion and word prediction on typing rates using on-screen keyboards. Assistive Technology, 18(2), 146–154.
[23]Pouplin, S., et al. (2014). Effect of a dynamic keyboard and word prediction systems on text input speed in patients with functional tetraplegia. Journal of Rehabilitation Research & Development, 51(3), 467–480.
[24]Jacob, R. J. K. (1990). What you look at is what you get: Eye movement-based interaction techniques. In Proceedings of the ACM International Conference on Human Factors in Computing Systems (pp. 11–18).
[25]Prabhu, V., & Prasad, G. (2011). Designing a virtual keyboard with multimodal access for people with disabilities. In Proceedings of the IEEE International Conference on Information and Communication Technologies (pp. 1133–1138).
[26]Singh, J. V., & Prasad, G. (2015). Enhancing an eye-tracker based human-computer interface with multi-modal accessibility applied for text entry. International Journal of Computers & Applications, 130(16), 16–22.
[27]Cretual, A., & Chaumette, F. (2001). Application of motion-based visual servoing to target tracking. The International Journal of Robotics Research, 20(2), 169–182.
[28]Gaskett, C., Fletcher, L., Zelinsky, A., et al. (2000). Reinforcement learning for visual servoing of a mobile robot. In Australian Conference on Robotics and Automation.
[29]Bustamante, G., Danes, P., Forgue, T., & Podlubne, A. (2016). Towards information-based feedback control for binaural active localization. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.
[30]Magassouba, A., Bertin, N., & Chaumette, F. (2018). Aural servo: sensor-based control from robot audition. IEEE Transactions on Robotics, 34(1), 169–186.
[31]Ghadirzadeh, S., Butepage, J., Maki, A., Kragic, D., & Bj orkman, M. (2016). A sensorimotor reinforcement learning framework for physical Human-Robot Interaction. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2682–2688).
[32]Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., & Hagita, N. (2006). Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning. Journal of Robotic Systems, 23(10), 545–554.
[33]Thomaz, A. L., Hoffman, G., & Breazeal, C. (2006). Reinforcement learning with human teachers: Understanding how people want to teach robots. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 352–357).
[34]Cruz, F., Parisi, G. I., Twiefel, J., & Wermter, S. (2016). Multi-modal integration of dynamic audiovisual

patterns for an interactive reinforcement learning scenario. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 759–766).
[35] Rothbucher, M., Denk, C., & Diepold, K. (2012). Robotic gaze control using reinforcement learning. In Proceedings of the IEEE International Symposium on Haptic Audio-Visual Environments and Games.
[36] Qureshi, A. H., Nakamura, Y., Yoshikawa, Y., & Ishiguro, H. (2016). Robot gains social intelligence through multimodal deep reinforcement learning. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 745–751).
[37] Qureshi, A. H., Nakamura, Y., Yoshikawa, Y., & Ishiguro, H. (2017). Show, attend and interact: Perceivable human-robot social interaction through neural attention Q-network. In Proceedings of the IEEE International Conference on Robotics and Automation.
[38] Vazquez, M., Steinfeld, A., & Hudson, S. E. (2016). Maintaining awareness of the focus of attention of a conversation: A robot-centric reinforcement learning approach. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.
[39] Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005). Towards a humanoid museum guide robot that interacts with multiple persons. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 418–423).
[40] Ban, Y., Alameda-Pineda, X., Badeig, F., Ba, S., & Horaud, R. (2017). Tracking a varying number of people with a visually-controlled robotic head. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.
[41] Yun, S.-S. (2017). A gaze control of socially interactive robots in multiple-person interaction. Robotica, 35(11), 2122–2138.