Assessment of Cognitive Load from Robot Assisted Sign Language Training

Ayca Aygun

December 16, 2021

In this paper, we explore the influence of learning sign language on cognitive load based on a sign language tutoring game between a human and a humanoid robot. The game includes several words chosen from American Sign Language (ASL) which are imitated by the robot using head gestures and body movements. During the experiment, similar words and gestures are imitated by the robot to intensify participant's mental load. To assess the cognitive load, the hand and body gestures of the participant, the response rates, the percentage of correct answers, and the response duration are examined. The results indicate that challenging questions, which are taken from similar words and gestures, result in increased number of hand and body gestures and decreased response performance while the response times are directly correlated with the level of difficulty. Moreover, the imitation of gestures with the robot have a positive effect on memorizing the signs. Additionally, the participants indicate better performance in practice part, which includes assembling freshly acquired knowledge.

1 Introduction

World Health Organization (WHO) reports that there are more than 430 million people globally who experience some kind of hearing disabilities which is more than 5% of the world's population [1]. Furthermore, over 700 million people will be expected to suffer from hearing loss by 2050 [1]. Sign language education is a significant practice to enhance the life quality of people who suffer from hearing disabilities. As reported by World Federation of the Deaf, nearly 70 million people worldwide use sign language as their primary language [4]. A well-designed education not only supports deaf people to gain knowledge but also it contributes them to construct a social interaction within their communities. There are several studies show that robots can be utilized for therapeutic

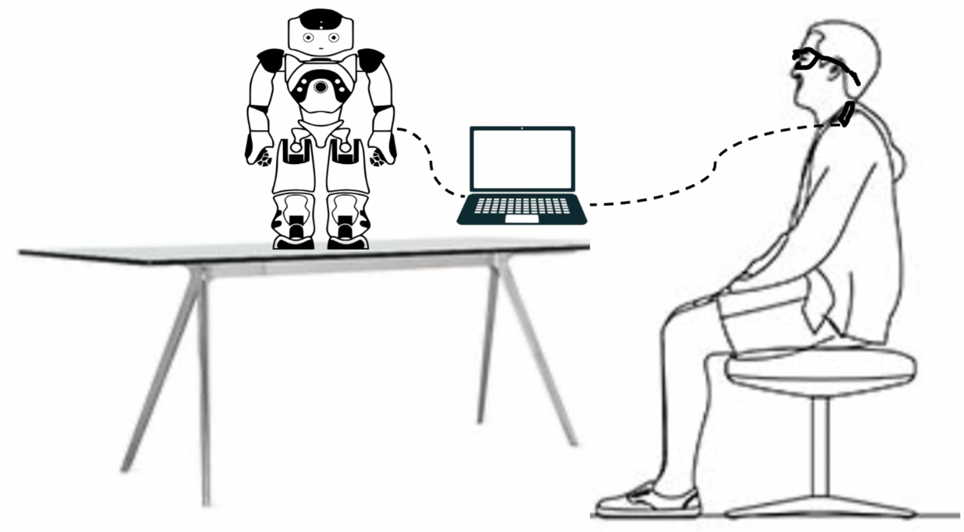

Figure 1: Experimental setup.

and educational needs in the field of human-robot interaction [2, 13]. Robot assisted education is advantageous due to several reasons. One of them is robot's efficacy at developing cognitive and perceptual outcomes [6]. Moreover, physical robots contribute moral development and language skills [23], and are capable of eliciting human's social attitudes which enhances learning performance [12]. The physical presence of robots in learning tasks is more worthwhile over virtual agents due to its positive and engaging impacts on humans [22]. Robot guided sign language tutoring, designed as an interactive learning environment, is an effective way of improving sign language skills of a human. To enhance the effectiveness of this methodology, the cognitive load of a participant should be carefully examined. In the course of learning, one way of assessing cognitive load is to explore participant's head and body gestures. One study indicates that processing load, which is needed to produce speech, is correlated with the number of gestures [16]. Another study proposes that gesture rate is higher in the case of less memorable and less perceptible tasks [15]. Testing learner's performance on memorizing the signs is another tool to evaluate the level of cognitive load. To test the response performance, the response rates over all of the questions and the correct answers over the responded questions are investigated. Response time is another indicator of response performance which may be expected to be directly correlated with the level of difficulty.

In this study, we design a robot assisted sign language tutoring game to investigate the effect of sign language learning on cognitive load. Figure 1 shows the experimental setup of the proposed method which includes a humanoid robot and a participant wearing an eye tracker. The tutoring includes two parts which are learning part and practice part. During the learning part, the humanoid robot imitates several sign language gestures consequently by mentioning their meanings and asks one of depicted signs to the participant in snatches. To increase the level of difficulty, which aims to generates higher cognitive load, we use either similar words or similar gestures. In the practice part, the robot imitates two signs, which were depicted before, and asks the participant to predict the possible meaning of the statement. This procedure targets to gain the mental load by forcing the person to assemble freshly acquired knowledge. The hand and body gestures along with the response duration, the rate of responses, and the rate of correct responses are investigated.

Our results indicate that the majority of the subjects hardly differentiate similar words and gestures which takes relatively longer response times. While responding the questions, participants tend to move their heads upwards or downwards as well as touching their heads. Some participants react the challenging questions by laughing or talking while the majority of the participants do not show hand or body movements in the presence of easier questions. The percentage of responses and correctly answered questions are decreased in the case of challenging questions. However, the response time indicates direct correlation with the the level of difficulty. Moreover, we observe that the imitation of gestures along with the robot results in a better comprehension of semantic meaning of signs.

The main contributions of this work are listed as follows:

- A sign language tutoring setting is designed as a game between a humanoid robot and a human to explore the mental load of a human participant caused by learning process.
- The level of difficulty is modified properly to generate higher mental load by leveraging similar words and gestures.
- The mental load of a participant is investigated by using multiple evaluation metrics such as the frequency of hand and body gestures, the response and correct answer rates among questions with various complexity levels, and response duration of the participants.

2 Background

There have been a few studies which investigate the dynamics of human-robot interaction in sign language education. One study proposes a robotic platform called RASA (Robot Assistant for Social Aims) which intents to guide children with hearing disorders to learn Persian Sign Language (PSL) from a humanoid robot [14]. The authors design a social robot for educational purposes which is capable of performing PSL. Although the study both presents the hardware design of the robotic platform and explores the sign language realization performance, the experiments are completed as an online survey where the robot displays different gestures from PSL. In our study, we conducted the experiments with the physical presence of both the humanoid robot and the human participant to raise the engagement and positive impacts between a robot and a human. Another study investigates the impacts of robotic sign language tutoring on children with autism spectrum disorder (ASD) [5]. The participants (children with ASD) are asked to mimic the gestures that the assistive robot depicts. The children, which completed the study with the attendance of their companions, indicate positive thoughts about the tutoring performed by the robot. Although this study demonstrates the efficiency of assistive robotics to keep children's attention with ASD disorder, the imitation of the signs used in this experiment is not effective due to robot's stiffness. An additional study proposes an interactive game between a humanoid robot which imitates the signs from Turkish Sign Language (TSL), and a human participant which learns the meanings of the signs and get tested [3]. They claim that a sign language tutoring with a humanoid robot is more effective way for children than video-based tutoring methods. The authors aim to augment the semantic meanings of the signs and validate the memorizing performance of the participants. To the best of our knowledge, this is the most relevant study to our method. Although the study provides a trustworthy experimental setup to assess the performance of humans on learning sign languages, the results are validated based only on the recognition rates of the participants. In our study, we evaluate the performance of the participants to estimate the cognitive load of the participants which is valuable to enhance the sign language tutoring settings. Moreover, we utilize additional metrics such as hand and body gestures, and human gaze to determine the mental load.

Several works have been proposed to explore the cognitive load during learning and memorizing processes [8, 20, 21]. Among those, only a few of them focus on assessing the mental load during sign language learning. One study investigates the impacts of captions on the mental load and the motivation of students with hearing disabilities [25]. The authors claim that deaf people gain and filter the knowledge with sign language which is their primary language. Hence, they suggest the simultaneous usage of sign language and captions to enhance the learning performance. The cognitive load is evaluated by means of scaling on different factors such as physical effort, mental effort, and task difficulty scored by the participants. In our study, we suggest using different metrics such as hand and body gestures, scattering of correct answers, and human gaze to estimate the cognitive load. Another research

There have been numerous of studies use human gaze to perceive people's cognitive states [9, 11, 17, 18, 24]. To assess the cognitive load of a human, a couple of eye gaze parameters are utilized such as blink rate, fixation rate, fixation duration, pupil dilation, and saccadic movements. One study introduce a driving simulator study to examine the capability of eye tracker to estimate the cognitive load of drivers [7]. The authors leverage from Detection Response Task (DRT) to reinforce the mental load and conclude that the pupil size and the blink rate increases with the increased secondary task difficulty.

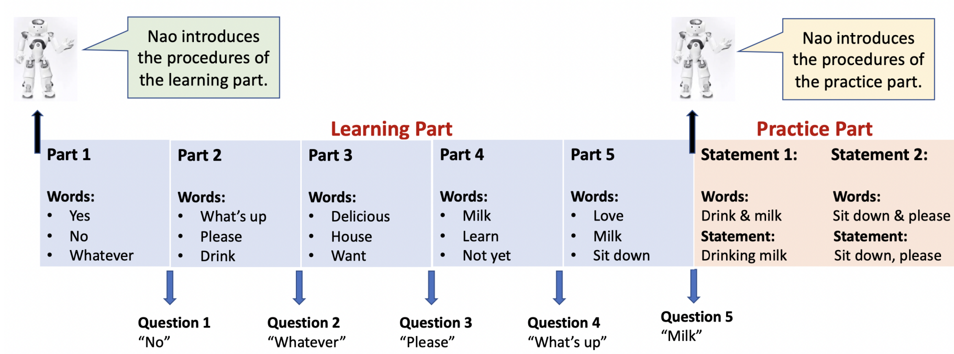

Figure 2: Design of the sign language learning tutoring.

3 Methodology

In this study, our aim is to design a sign language learning environment for people with hearing disorders by investigating cognitive load as a result of memorizing the signs one after another. However, the preliminary study was tested with people with no hearing problem to test our method.

3.1 Data Collection

The experiment includes a humanoid robot and a participant wearing an eye tracker to record the gaze data. The robot simulation and the eye tracker were initiated at the same time for simultaneous recording. The Nao was placed on the table in front of the participant sitting on a chair. The distance was set approximately one meters between the robot and the participant to avoid any possible accident. The participants were informed about the setup of the game and the approximate duration of the experiment.

3.2 Computational Approach

To evaluate the response times of the participants to the questions, we used three time scales for learning part (responding between $0-1$, $1-3$, or $3-5$ seconds and three time scales for practice part (responding between $0-1$, $1-5$, and $5-10$ seconds). We annotated the onset of the responses as the ending point of the question (the imitation of a gesture that robot performs). If the participant responded the question before the robot completed the imitation of the gesture, we assumed that the participant responded it within $0-1$ seconds. The response performances were evaluated for the learning and practice parts separately. The response percentage and the true response percentage were determined as the percentage of responses over the total number of questions and the percentage of correct answers over the total number of responded questions, respectively.

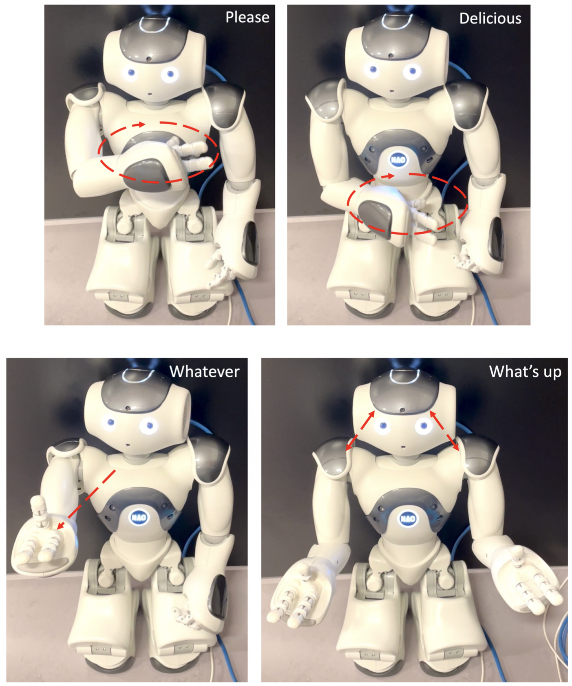

Figure 3: The robot shows two signs with similar gestures (upper) and two signs with similar words (bottom).

3.3 Hardware and Software Design

We used the humanoid robot Nao which was developed by SoftBank Robotics in 2008. It is 58cm in height with a 25 degrees of freedom and it has 7 touch sensors placed on the head, hands and feet. Nao is capable of interacting with humans with its four directional microphones and speakers to interact with humans. Nao was programmed by Choregraphe which is a graphical environment created by Aldebaran Robotics.

3.4 Experimental Design

Each participant performed approximately 6.30 minutes long sign language game depicted in Figure 2. The game included one learning part and one practice part. Learning part contained five sub-parts and within each sub-part, Nao imitated three gestures (each of them twice) taken from American Sign Language (ASL) while mentioning its meaning vocally. We selected hand and body gestures proper to the design of the robot by considering that Nao has three fingers at each hand which limits the imitation of some finger-based signs. At the end of each sub-part, Nao showed one gesture and asked the participant to tell the meaning of the sign within 5 seconds. The question was taken from the sign pool depicted by the robot from the beginning of the game (*e.g.*, the 4^{th} question was selected from the 2^{nd} sub-part). To do this, we aimed the participants to have gradually increased mental load through the end of the learning part. In the course of learning part, the robot depicted two similar gestures and two similar words to boost the cognitive load by raising the level of difficulty Figure 3 indicates the similar gestures and similar words. The upper left and the upper right pictures illustrate the robot to imitate "Please" (make a circle on the chest with one hand) and "Delicious" (make a circle on the belly with one hand), respectively. Moreover, the bottom left and the bottom right pictures represent the robot to imitate "Whatever" (open the palm and stretch the hand forward) and "What's up" (open both palms and shake the head up and down), respectively. There of the questions were selected from these four gestures. The practice part contained two questions. In each question, Nao showed two gestures and asked the participant to guess the meaning of the possible statement (*e.g.*, "Sit Down" and "Please" generates the statement "Sit down, please). We provided 10 seconds to the participants to answer the practice questions. The goal of the practice session was to boost the mental load of the participants to make them to fuse the freshly gained knowledge.

3.5 Measures

We evaluated the performance of the study in two categories. First, the hand and body gestures were examined in terms of the frequency of different body movements and reactions. Second, we evaluate the response performance of participants by calculating the rate of responses over all questions, the rate of correct answers over the responded questions, and the response times. Moreover, we investigated additional individual situations such as the question which was answered by only one participant or the participant who correctly answered the word in practice part while could not answer the same word within learning part.

3.6 Participants

The study was performed by five volunteers who are graduate students at Tufts University. The participants included 2 females and 3 males with no preliminary knowledge of sign language and no hearing disorders. English is the native language of 60% of the participants. Figure 4 shows a participant playing the game.

4 Results

We evaluated the performance of our method in three main categories: The hand and body gestures of the participants, the response performances such as the rate of correct answers and the response times, and the blink rates during

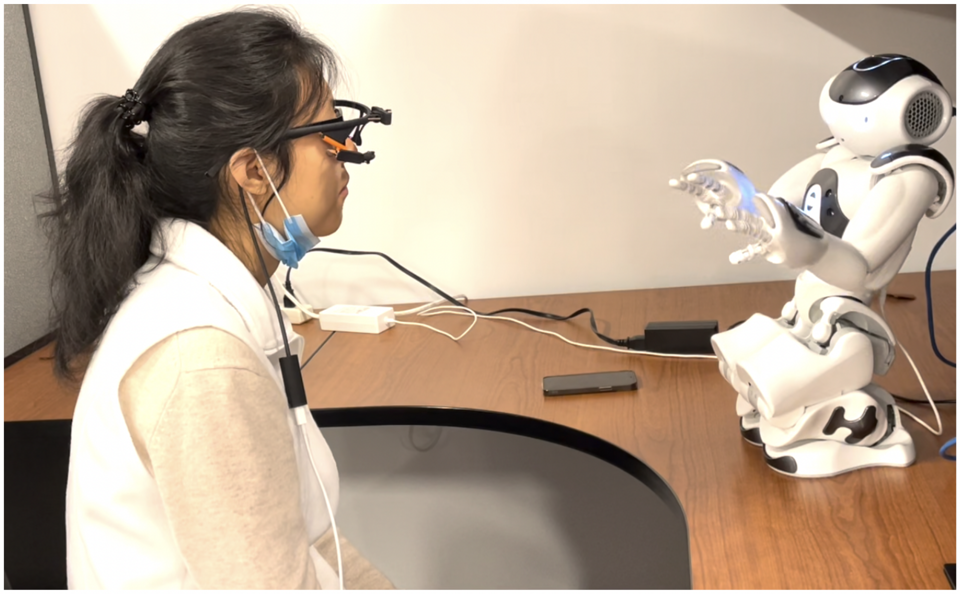

Figure 4: Participant performs the game with the robot.

the answering process. The results indicate that the level of difficulty results in increased number of head/body movements, longer response times, and decreased number of correct answers. Moreover, the participants tend to blink more frequently in the case of challenging questions.

4.1 Hand and Body Gestures

Table 1: Hand and body gestures for different questions

Questions / Gestures	"No"	"Whatever"	"Delicious"	"What's Up"	"Milk"
Thinking loudly	×	✓	×	✓	×
Laughing	×	✓	✓	×	✓
Use of "uhm/umm"	×	×	✓	✓	×
Look around	×	✓	✓	✓	×

We investigate the scattering of hand and body movements during the learning part. Four gestures are observed which are "Thinking loudly/talking", "Laughing", "Use of uhm/umm", and "Look around/upward/downward" indicated in Table 1. The body movements are marked as checked if it was generated by at least one participant. The results indicate that participants tend to react more to the similar words/gestures which are assumed to create higher cognitive load. For example, three different gestures were observed from at least one participant in the case of the similar words "Whatever" and "What's up", and "Delicious" which represents a similar gesture with "Please". It demonstrates that higher cognitive load is a possible cause of increased number of head and body gestures.

4.2 Response Performance

The response performance of the participants was evaluated with three metrics. First, the percentage of the responses over all questions and the percentages of correct responses over the responded questions were calculated for learning and practice parts separately. Second, we evaluate the response times in terms of easier and harder questions. Third, we have done additional performance analysis for specific conditions such as the only correct response to "Whatever".

4.2.1 True/False Responses

Table 2 shows the responded question rate over all questions and correctly answered response rate over responded questions. The results indicated that the response rate is lower for tricky questions taken from similar words and gestures such as "Whatever", "Delicious", and "What's up" compared to the easier questions. Moreover, the lower performance of true responses was observed as 20% for "Whatever". The performance of the participants was better for practice part relative to the learning part. The usage of two gestures together may encourage the participants to assemble freshly gained knowledge which result in a better memorizing performance.

Table 2: Percentages of Responded Questions and Correct Responses

Metric \ Questions	"No"	"Whatever"	"Delicious"	"What's Up"	"Milk"	"Practice 1"	Practice 2
Responded Questions	100	60	100	80	80	100	100
Correct Responses	100	33.3	40	50	100	100	75

4.2.2 Response Times

Table 3 indicates the scales of response times calculated from responded questions. The response times are categorized by using three scales. The response duration is comparatively higher for more challenging questions such as "Whatever" and "What's up" than for the easier questions such as "No" and "Milk". Although the performance of "Delicious" seems higher, only 40% of the participants could answer this question truly. The results demonstrate that there is an inverse correlation between the difficulty level of the questions and the response times.

Table 3: Scales of Response Times for Learning Part

Metric \ Questions	"No"	"Whatever"	"Delicious"	"What's Up"	"Milk"
0-1 seconds	100	-	100	25	100
1-3 seconds	-	33.3	-	25	-
3-5 seconds	-	66.6	-	50	-

Moreover, Table 4 depicts the scales of response times over all responded questions in practice part. The evaluation was performed or four participants as one participant could not complete the practice part due to technical issues. The results show that the majority of the participants responded the practice questions within 5 seconds. Moreover, 75% of the participants answered the questions less than one second.

Table 4: Scales of Response Times for Practice Part

Metric \ Questions	"Practice 1"	"Practice 2"
0-1 seconds	75	75
1-5 seconds	25	25
5-10 seconds	-	-

4.2.3 Additional Observations

We observed that the only participant, who responded the word "Whatever" correctly, mimic the gesture along with the robot. Figure 5 shows the participant with replicating the gesture represents "Whatever" with the robot. It demonstrates that imitation of signs would contribute better comprehension of semantic meanings of the signs. Moreover, one participant successfully responded the statement "Drink milk" in practice part after a wrong answer to "Milk" in learning part. This indicates the advantage of using statements on refreshing the memory.

5 Discussion

This study shows that the frequency of participant's body gestures have a direct relationship with the complication level of the questions. Participants react the challenging questions with several body movements such as looking upward/downward, laughing, talking, and using words such as "uhm/umm". The response rates and the percentage of truly answered questions are increased for the less challenging questions. However, the response times have a direct

Figure 5: The participant replicate the gesture with the robot.

correlation with the level of difficulty (*i.e.*, participants have longer response duration for hard questions selected from similar words and gestures). Thus, the mental load leads to increased number of hand and body gestures as well as worse response performance in terms of response rates, correctly answered question percentages, and response duration.

During the learning part, the question "Whatever" was correctly answered only by one participant. It is observed that the participant, who responded the question truly, was the only person who imitated the gesture along with the robot. This demonstrates that the imitation of the gestures would lead better comprehension of the semantic meanings of the words. Moreover, one participant, who answered the question "Milk" wrongly, could successfully answer the statement of "Drink milk" in the practice part. This indicates that the usage of assembling words as a statement would provide the participants to refresh their memory and remember the meaning of the words.

In the future work, human gaze can be used to investigate the mental status of a participant. Spontaneous blink rate is one of the main human gaze parameters which is related to dopaminergic activity in the brain. Dopaminergic activity is generated by a significant neurotransmitter called by dopamine which is associated with learning and goal-oriented attitude [10]. Fixation rate is another significant human gaze parameter to estimate mental load. In [19], number of fixation is linked to the field of dependence-independence where field-dependent people have struggle with identifying details and field-independent people pay attention to details. In this regard, a careful examination of human gaze contributes better comprehension of cognitive load.

Additionally, the study may be designed more interactive with a proper implementation of humanoid robot. To achieve this, the robot can be programmed to adjust the difficulty level of the questions with respect to the performance of the participant. For example, the robot asks less challenging questions in the case of worse response performance which may be advantageous to keep participant's interest to the game. This procedure will require image recognition property at robot's side.

6 Conclusion

In this paper, we design a sign language tutoring as a game between a humanoid robot and a human participant to assess the cognitive load of a human. The robot is programmed to imitate several sign gestures and asks questions from depicted gestures to the human participant. To generate a mental load, similar words and gestures are leveraged in the course of training. The mental load is evaluated with the frequency of hand and body movements, the response rates, the percentage of correctly answered questions, and response duration. The results indicate that the cognitive load is correlated with the number of gestures and response duration. Moreover, the rate of responses and the percentage of correctly answered question have opposite relationship between the level of difficulty. In the future, the human gaze parameters such as blink rate and fixation rate may be utilized for the better assessment of cognitive load. Moreover, the efficiency of the design can be improved by programming the robot more interactive.

References

[1] Deafness and hearing loss. `https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss`. Accessed: 2021-10-01.

[2] Aggarwal, S., Gupta, D., and Saini, S. (2019). A literature survey on robotics in healthcare. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pages 55–58. IEEE.

[3] Akalin, N., Uluer, P., Kose, H., and Ince, G. (2013). Humanoid robots communication with participants using sign language: An interaction based sign language game. In *2013 IEEE workshop on advanced robotics and its social impacts*, pages 181–186. IEEE.

[4] Amor, A. B. H., Ghoul, O., and Jemni, M. (2017). Toward sign language handshapes recognition using myo armband. In *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, pages 1–6. IEEE.

[5] Axelsson, M., Racca, M., Weir, D., and Kyrki, V. (2019). A participatory design process of a robotic tutor of assistive sign language for children with

autism. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8. IEEE.

[6] Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, 3(21).

[7] Čegovnik, T., Stojmenova, K., Jakus, G., and Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Applied ergonomics*, 68:1–11.

[8] Choi, H.-H., Van Merriënboer, J. J., and Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review*, 26(2):225–244.

[9] Ikehara, C. S. and Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. In *Proceedings of the 38th annual hawaii international conference on system sciences*, pages 295a–295a. IEEE.

[10] Jongkees, B. J. and Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—a review. *Neuroscience & Biobehavioral Reviews*, 71:58–82.

[11] Jyotsna, C. and Amudha, J. (2018). Eye gaze as an indicator for stress level analysis in students. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1588–1593. IEEE.

[12] Kennedy, J., Baxter, P., and Belpaeme, T. (2015). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, 7(2):293–308.

[13] Krebs, H. I. and Hogan, N. (2006). Therapeutic robotics: A technology push. *Proceedings of the IEEE*, 94(9):1727–1738.

[14] Meghdari, A., Alemi, M., Zakipour, M., and Kashanian, S. A. (2019). Design and realization of a sign language educational humanoid robot. *J. Intell. Robotic Syst.*, 95(1):3–17.

[15] Melinger, A. and Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and cognitive processes*, 22(4):473–500.

[16] Morsella, E. and Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American journal of psychology*, pages 411–424.

[17] Prabhakar, G., Mukhopadhyay, A., Murthy, L., Modiksha, M., Sachin, D., and Biswas, P. (2020). Cognitive load estimation using ocular parameters in automotive. *Transportation Engineering*, 2:100008.

[18] Rafiqi, S., Wangwiwattana, C., Kim, J., Fernandez, E., Nair, S., and Larson, E. C. (2015). Pupilware: towards pervasive cognitive load measurement using commodity devices. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8.

[19] Raptis, G. E., Katsini, C., Belk, M., Fidas, C., Samaras, G., and Avouris, N. (2017). Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies. In *proceedings of the 25th conference on user modeling, Adaptation and Personalization*, pages 164–173.

[20] Skulmowski, A. and Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. *Frontiers in psychology*, 8:1191.

[21] Verhoeven, L., Schnotz, W., and Paas, F. (2009). Cognitive load in interactive knowledge construction.

[22] Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 872–877. IEEE.

[23] Wei, C.-W., Hung, I., et al. (2011). A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *Turkish Online Journal of Educational Technology-TOJET*, 10(2):11–23.

[24] Wolf, E., Martinez, M., Roitberg, A., Stiefelhagen, R., and Deml, B. (2018). Estimating mental load in passive and active tasks from pupil and gaze changes using bayesian surprise. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, pages 1–6.

[25] Yoon, J.-O. and Kim, M. (2011). The effects of captions on deaf students' content comprehension, cognitive load, and motivation in online learning. *American annals of the deaf*, 156(3):283–289.