10.3217/978-3-99161-014-4-008

A NOVEL CHATGPT-DRIVEN COMMUNICATION AID BASED ON CODE-MODULATED VISUAL EVOKED POTENTIALS (CVEP)

Atilla Cantürk, Ivan Volosyak

Faculty of Technology and Bionics, Rhine-Waal University of Applied Sciences, Kleve, Germany

E-mail: ivan.volosyak@hochschule-rhein-waal.de

ABSTRACT: Brain-computer interface (BCI) systems, including applications based on visual evoked potentials (VEPs), have proven to provide reliable and accurate control. In recent years, communication has remained one of the main application areas of modern BCIs, with a lot of advancements based e.g., on the incorporation of dictionary support and text prediction. This study explores the integration of BCIs with artificial intelligence (AI), specifically focusing on the development and evaluation of an innovative spelling interface powered by the ChatGPT application programming interface (API). Aimed at enhancing communication for individuals with severe motor impairments, this interface combines the precision of code-modulated visual evoked potentials (cVEPs) with the predictive capabilities of AI to offer a more intuitive and efficient user experience. The performance of 13 healthy participants (10 females) was evaluated in an online experiment. The participants successfully completed all spelling tasks using the cVEP BCI with aid from ChatGPT, achieving a mean information transfer rate (ITR) of 33.16 bpm, a mean accuracy of 87.49%, and an average output of 8.74 output characters per minute (OCM) for unique sentence tasks. This was slower than in our previous research using an n-gram model which achieved 18.9 characters per minute.

INTRODUCTION

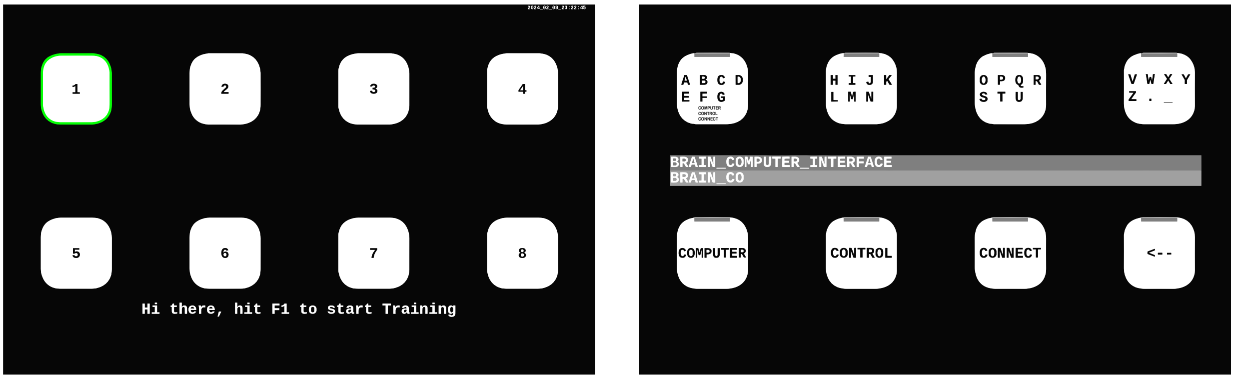
A BCI system detects, analyzes and decodes brain activity in real time to provide communication with the external environment, without involving normal output pathways of the human nervous system or muscle activities [1]. Modern BCIs can be used as communication tools for severely impaired people suffering for example from spinal cord injuries, brain stem strokes, amyotrophic lateral sclerosis (ALS), or muscular dystrophies. For the practical use of such BCI applications their accuracy and speed are the most important factors. Visual evoked potentials (VEPs) allow the fastest BCI realisation, between them the code-modulated VEPs (cVEPs), where all stimuli are modulated with different time lags of the same code sequences, typically the m-sequences, yield potentially higher accuracies [2]. Further increase in the communication speed, usually measured in terms of information transfer rate (ITR) is possible by using word prediction methods, particularly based on n-gram models. In our previous paper [3], c-VEP based BCI system was further extended by several methods for enhanced target identification, including dynamic sliding windows and software-based stimulus synchronization, coupled with an ensemble-based classification. Integrating a dictionary-driven n-gram word prediction model, the system demonstrated improved usability, with significantly better results when the dictionary integration was used. Unfortunately, this software implementation of the signal processing and the dictionary support in form of a single custom-made application limits the transfer of this code into newer BCI applications, necessitating the re-development of the dictionary support. Another issue is that until recently, it was not technologically feasible to accurately reproduce sentences generated using such communication aids, especially concerning the proper verb declension, which requires a complete understanding of the sentence's information content. However, with the advent of new AI-based language systems, it is now possible to bridge this gap and produce stylistically and grammatically correct sentences.

In recent years, AI has garnered significant attention across various domains, revolutionizing the way we interact with technology and transforming traditional workflows. From chatbots for libraries [4] to recommendation system for farms [5], AI-driven solutions have demonstrated remarkable capabilities in understanding and processing human language. The combination of BCIs and AI marks a significant shift in human-computer interaction, especially for individuals with severe motor impairments. Advanced AI models like ChatGPT enhance this synergy, revolutionizing interactions from healthcare to customer service. Its applications, aiding clinical diagnoses to supporting medical education, highlight AI's utility, though ethical and legal considerations accompany it [6]. The introduction of ChatGPT has sparked a robust debate over its potential applications and limitations, underscoring the need for a nuanced exploration of AI's role in healthcare and medical research.

This paper presents an innovative spelling interface that leverages the ChatGPT API, demonstrating the seamless integration of cVEP-based BCIs with AI to create a more intuitive and efficient communication tool. By examining the advantages, limitations, and effects of employing ChatGPT and AI in such interfaces, alongside their practical applications and future prospects in medicine and

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-008



(a) "Training" mode, awaiting initiation.

(b) Online "Two-Steps Speller" interface with dictionary hints (top right).

Figure 1: Fig. 1a shows 1st selection field highlighted in green and waiting for "F1" key press to start the training session. Once training begins, all text outside of the selection fields will be hidden. In Fig. 1b, API word suggestions appear in selection fields 5 to 7, with dictionary hints also shown in the 1st selection field (placed inside the field, under the letter groups).

healthcare, we aim to contribute to the ongoing discourse on the responsible and effective use of AI technologies in enhancing human-computer interaction and healthcare outcomes.

MATERIALS AND METHODS

Participants: 13 participants (10 females) participated in this study; the average age of the subjects was 24.46 years, with a standard deviation (SD) of ± 4.2. All participants provided written consent in adherence to the Declaration of Helsinki, and the study received approval from the ethical committee of the medical faculty at the University Duisburg-Essen. The collected data for analysis purposes were stored anonymously, ensuring the confidentiality of the participants. Subjects received compensation for their involvement in our study. Since this was not addressed in the ethical approval, the EEG data cannot be published.

Hardware: The computer in use was a Dell Precision Desktop with NVIDIA RTX3070 graphics card that operated on Microsoft Windows 10 (21H2) Education running on an Intel processor i9-10900K (3.70 GHz). For the purpose of presenting the stimuli, a modern display (Asus ROG Swift PG258Q, Full-HD, 240 Hz maximal refresh rate) was used.

An EEG amplifier (g.USBamp medical engineering GmbH, Schiedlberg, Austria) was used, utilizing all 16 signal channels which were placed according to the international system of EEG electrode placement at positions: P7, P3, Pz, P4, P7, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, O9, Iz and O10. Additionally, the reference electrode was positioned at Cz, while the ground electrode was placed at AFz. During the preparation stage, regular abrasive electrolytic electrode gel was used between the electrodes and the scalp to reduce impedances to less than 5 kΩ.

GUI: The graphical user interface (GUI) is illustrated in Fig. 1b. An eight target spelling interface as presented in [3] was utilized. Selecting individual characters required exactly two steps ("two-steps speller"). The graphical user interface (GUI) was designed with its first row featuring 28 characters, including the 26 letters of the alphabet, an underscore, and a full-stop character, organized into four boxes, each containing seven characters. The second row provided three suggestions generated by the ChatGPT API (dictionary suggestion boxes), along with an option for correction. Utilizing the correction option allowed users to delete the last typed character or word, enhancing the typing experience by integrating both predictive text and error correction functionalities. By selecting a letter group from the first row, the associated characters were presented individually in the "second step". The GUI includes dictionary hints that present the same recommendations as the ChatGPT API at the bottom of the lastly selected letter, facilitating easier selection without the need to divert attention. An illustration of this feature is provided in Fig. 1b (top-left corner selection field). Each selection triggers audio and visual feedback (the selected field briefly enlarges and turns green for correct selections, or turns red for incorrect selections).

Stimulus Presentation: The stimuli targeted in the experiment were comprised of eight selection fields(boxes) (230×230 pixel) arranged as 2×4 selection field matrix (see Fig. 1). 63 bit m-sequences c_i, $i = 1, \dots, K$ ($K = 8$ for our case) were assigned to the selection field matrix employing a circular shift of 4 bits (c_1 had no shift, c_2 was shifted by 4 bits to the left, c_3 was shifted by 8 bits, etc.). The codes were allocated to the matrix in a row-wise manner, beginning with the upper left target labeled as c_1, and subsequent targets were labeled following a row-major sequence. The stimuli linked to the codes switched between "black" (the background color, denoted by "0") and "white" (indicated by "1"). Here, c_1 was defined as

$$\begin{aligned} c_1 = {} & 10101100110111011010010011100010 \\ & 1111001010001100001000001111110 \end{aligned} \tag{1}$$

The duration of a stimulus cycle in seconds can be calculated by dividing the code length by the monitor refresh

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-008

rate r in Hz; in this experiment, $63/60 = 1.05s$ (the used refresh rate was 240 Hz, so the stimulus changed in accordance with the bit sequence, but for every fourth frame). Spatial filters were developed using the information gathered during the recording phase for classification. Canonical correlation analysis (CCA) was used on the training trials in this regard. Further details about used cVEP signal processing methods can be found in [2].

Training: During the recording phase, eight stimuli were observed sequentially from 1 to 8 by the participants, as illustrated in Fig. 1a. The recording was divided into six blocks of training, denoted as $n_b = 6$. Within each block, every stimulus was focused on once, resulting in a total of $6 * 8 = 48$ trials. Each trial lasted for 2.1 seconds, during which the code pattern was displayed for two cycles. A visual cue, represented by a green frame, indicated the specific box towards which participants were required to direct their gaze. Following each trial, the subsequent field the user needed to focus on was highlighted, and the flickering paused for one second. After completing each block of eight trials (all eight targets), the software transitioned to the next block of training, with a one-second pause until 48 trials were accomplished.

ChatGPT API: The ChatGPT application programming interface (API) works by sending a prompt, typically a piece of text or a question, to the API endpoint. The model then processes this input and generates a response that continues the conversation or provides relevant information based on the context provided in the prompt.

In order to be able to construct API requests, we modified our C++ based software with the help of the documentation provided by OpenAI [7]. We used libcurl to enable our software to be able to communicate with the API and used a JSON parser to make it easier to work with the response from the API. We made adjustments to certain parameters to facilitate the use of the API. For detailed information on the parameters modified and their specific functions, please refer to Table 1.

The word completion algorithm functions by identifying words demarcated by underscore (“_") characters. When a user selects a suggested word via the dictionary buttons (see Fig. 1b), the algorithm updates the text by replacing the characters entered after the most recent underscore character with the selected suggestion. When the user selects the correction button after choosing a suggested word, the software restores their manually typed text to its original form before the suggestion was applied.

Once user types a letter, the typed sentence is added into the API request and sent. Once the response is received the contents are extracted and separated into 3 different words and pushed into the dictionary suggestion boxes and also into dictionary hints text, giving the user the chance to type the word recommended by ChatGPT.

Experimental Protocol: The experiment was conducted in the BCI-Lab of Rhine-Waal University of Applied Sciences (HSRW). Firstly participants filled a questionnaire with questions regarding their experience with

Table 1: Key ChatGPT API Parameters

Model: “gpt-3.5-turbo-0125”. Chosen for its low cost and high speed.
Instruction: “We are trying to realise the speller. Always return suggestions for just the most likely last word of the query starting with letters of the query, having in mind previous words of the query. Do not return previous words of the query. Always return only three comma-separated words, no further information, no line skips." Customized for spelling suggestions (52 tokens). This parameter guides the model on how to generate its response depending on the input prompt.
Max tokens: Set to 100. Max tokens parameter represents the maximum number of tokens that can be generated in the chat completion.
Presence_penalty and **Frequency_penalty:** Set to -0.1 to fine-tune response variability. Both parameters slightly increase the likelihood of repeating information, promoting a less random output.
Top_p: Adjusted to 0.1 for focused response generation. This parameter controls the diversity of the model’s responses by limiting the probability mass considered for sampling the next token. Setting it to 0.1 ensures that the model’s outputs are highly focused and relevant to the given prompt, by choosing the response from the top 10%.

BCI systems. Participants were also asked to provide insights into their experiences with the BCI technology and describe their level of fatigue prior to initiating the study; their answers were recorded. Then, participants were briefed about the procedure and the operation of the speller. Following these explanations, participants engaged in a preliminary test run to accustom themselves to the speller, during which they freely composed a sentence of their choice and got familiar to the use of ChatGPT word recommendations. The threshold, gaze shift, and time window settings were calibrated as necessary during the familiarization. During this study, participants were told to spell the words “BCI”, “KLEVE” then spelled the pangram “THE_QUICK_BROWN_FOX_JUMPS_OVER_THE_LAZY_DOG”. After successfully finishing the spelling of the pangram, a unique sentence for each participant was randomly chosen from a pool of sentences that were inspired by news article titles.

After successfully completing the spelling session, participants completed the post-questionnaire containing questions regarding their impressions, opinions and their experience towards the BCI systems. The questions regarding their experience composed of questions regarding the flickering lights and how it affected them and questions regarding the effect of the assistance from the API and the dictionary hints functionality.

Spelling phases concluded automatically upon correct word spelling. On average, each subject’s spelling session (just spelling) lasted 20 to 25 minutes. Resulting accuracy, ITR, and OCM values were recorded for all com-

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-008

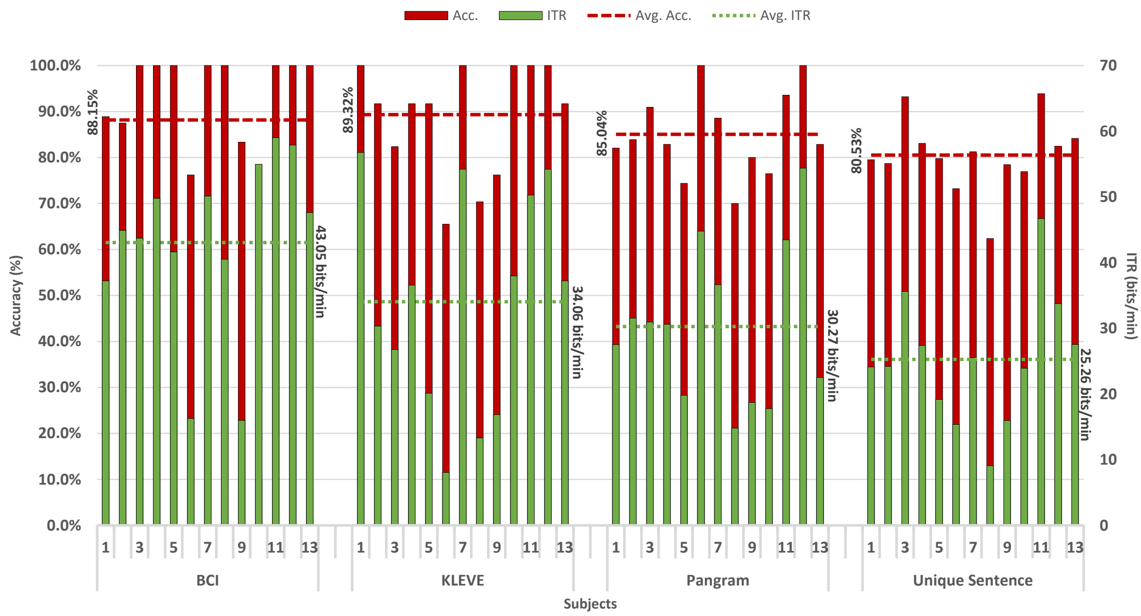


Figure 2: Individual spelling task's accuracy, Information Transfer Rate (ITR) and their average values are presented in sequence. Participants were assigned the following typing tasks:"BCI", "KLEVE" "THE_QUICK_BROWN_FOX_JUMPS_OVER_THE_LAZY_DOG" and a unique sentence for each participant.

pleted tasks.

Evaluation Measures of the BCI Performance: The BCI system's performance was evaluated using commonly used accuracy (Acc.), Information Transfer Rate (ITR), and in form of the output characters per minute (OCM).

Accuracy: The accuracy was calculated by dividing the total number of correct selections (word completions were considered a single command), including user-necessary corrections during speller execution, by the overall commands classified. The resulting accuracy value was displayed as a percentage value on the speller interface.

ITR: The Information Transfer Rate (ITR) was calculated in bits per minute (bits/min) using the formula:

$$B = \log_2 N + P \log_2 P + (1-P)\log_2\left[\frac{1-P}{N-1}\right], \qquad (2)$$

where:
B= information transferred in bits,
N= number of targets (for this study it is equal to 8),
P= classification accuracy.

To obtain the ITR in bits/min, B is multiplied by the average classification time in minutes. For more information and tools to calculate ITR, visit our webpage: `https://bci-lab.hochschule-rhein-waal.de/en/itr.html`.

OCM: The Output Characters per Minute (OCM) measures typing speed by dividing the total number of output characters by the time taken to type them. OCM accounts for error correction time, as participants will require additional time for corrections if mistakes are made.

Evaluation of the Questionnaire: A questionnaire was designed to collect participant feedback, with sections dedicated to both pre-experiment and post-experiment questions. These sections are intended to be completed respectively before and after the experiment, focusing on assessing user experience and the improvements ChatGPT has made to user comfort. For further information, refer to Table 2, which outlines these pre- and post-experiment questions.

RESULTS

The results indicating BCI performance are shown in the Fig. 2 and Fig. 4. Fig 2 illustrates the ITR and accuracy values achieved by participants, along with average values per task. As tasks lengthened, average ITR decreased, and sentence accuracies were lower compared to single-word tasks. Fig 4 displays the output characters per minute (OCM) values and their averages per task, revealing that the average OCM values are lower for the first two tasks compared to the last two. This difference is attributed to ChatGPT's inability to predict these words, likely due to their uncommon usage. However, ChatGPT notably enhanced participants' performance in the final two tasks compared to its performance in previous tasks. This was especially noticeable in the pangram task, where the average Output Characters per Minute (OCM) increased dramatically, more than tripling from 5.72 characters per minute to 17.72 characters per minute, having in mind the total numbers of spelled characters.

Results from the questionnaires indicate that eight out of 13 participants felt more tired after the experiment, while the rest reported no change in their fatigue levels. Four out of 13 considered the flickering disturbing. Majority

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-008

Table 2: Used Questionnaires.

Pre-Questionnaire
Have you ever used a BCI system? If yes, please add some information about it.
Do you have a vision prescription? If yes, are you wearing a reading aid now?
How tired do you feel right now? 1: not at all, 6: very much
How many hours did you sleep last night?
Post-Questionnaire
How tired do you feel right now? 1: not at all, 6: very much
Did you find the flickering disturbing? 1: not at all, 6: very much
Was it easy for you to concentrate on the boxes? 1: not at all, 6: very much
Did dictionary hints improve word completion? 1: not at all, 6: very much
Do you prefer the speller with or without ChatGPT aid? With / Without
Would you repeat the experiment? Yes / No / Maybe
Could you use the system daily? Yes / No / Maybe
In your opinion, how long can the system be used without breaks?
What was the unique sentence you had to type?
Do you think the BCI is a reliable control method? Yes / No / Maybe

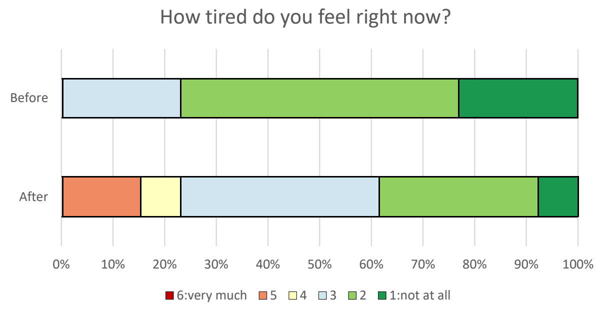


Figure 3: Participant feedback on fatigue levels before and after the experiment, measured on a scale from 1 (not fatigued at all) to 6 (very much fatigued).

found concentrating on the boxes easy. Everyone preferred spelling with the aid of ChatGPT and 12 out of 13 found dictionary hints helpful. 12 out of 13 participants reported that they would like to take the experiment again. On average, participants reported that they could use the system for approx. 1.2 hours. Majority of the participants found dictionary hints helpful. Findings related to fatigue levels are presented in Figure 3. Eight participants had no vision prescription, two had prescriptions but opted not to use any corrective wear, and the remainder used their prescribed vision aids.

DISCUSSION

The general use of digital technologies owned by private companies and located overseas raise many data protection, ethical, safety and security questions. E.g., it is well known that OpenAI has recently removed accounts of hacker groups from China, Russia, North Korea, and Iran. The use of ChatGPT as a language model for the BCI purposes is of course not comparable to this example, but, on the other hand, texts produced by the target group of users with disabilities, a most vulnerable group, need much more privacy and require careful ethical considerations.

We encountered several challenges during the experiments and development phases. One notable issue was ChatGPT's occasional difficulty in accurately following instructions, leading to incorrect recommendations. For instance, in some instances, instead of generating the complete word as instructed, it would only output the missing part of the word, despite clear instructions to do the opposite. We tested many different instructions to minimize this issue, but there might be of course a better instruction set. Using a different ChatGPT version will also likely require a completely different set of queries. We fine-tuned several supplementary parameters (see Tab. 1), to refine and control the API's behavior. We should note that although ChatGPT occasionally failed to adhere strictly to instructions, it sometimes succeeded in enhancing typing speed by suggesting corrections and even predicting the next word before the user began typing it. This behavior became evident to us during some instances when participants were reciting pangrams (see Fig. 4), where the OCM values were the highest. Additionally, it demonstrated the ability to switch languages seamlessly, without needing explicit commands (for instance, recommending German words upon typing "KLEVE"), while still managing to follow instructions to a satisfactory extent. Another challenge encountered was occasional unresponsiveness of the software due to high network traffic impacting the ChatGPT, leading to delays while awaiting responses. For best performance it is recommended to have a good network connection and the API status should be checked (`https://status.openai.com/`). Following an extended period of time spent attempting to type the initial words "BCI" and "KLEVE", two participants opted to withdraw from the experiment (therefore, in total 15 participants were recruited for this study). Some factors that affected the performance include the frequent need for words to be in plural form or to have different endings, which required additional typing for ChatGPT to suggest the appropriate word forms. Another common issue was participants ignoring suggestions and opting for manual typing, resulting in lower characters per minute. When comparing our average OCM with [8], which used an n-gram prediction model, a clear difference is observed. Specifically, for sentences aided by ChatGPT, the average OCM

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-008

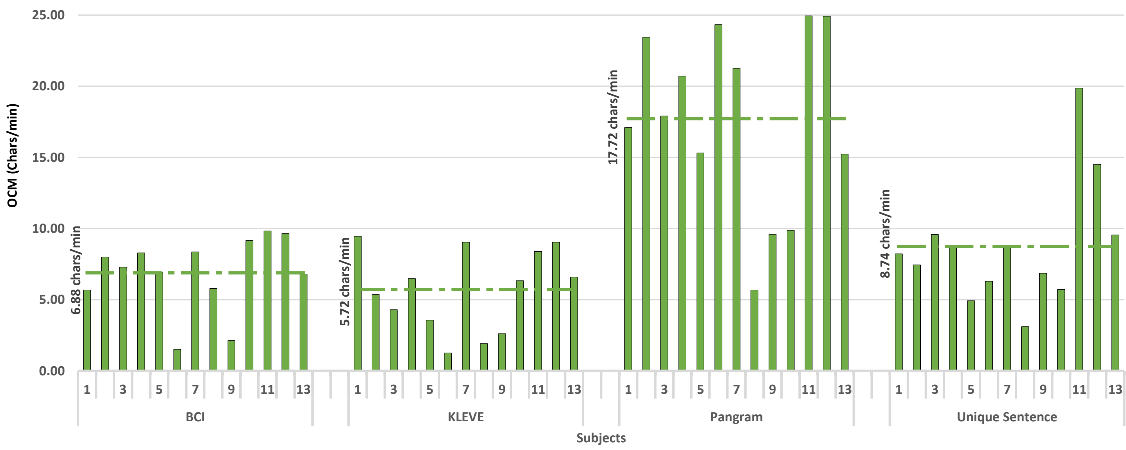


Figure 4: For each task, the Output Characters per Minute (OCM) are represented by bars, and the average values are illustrated with lines. Participants were assigned the following typing tasks:"BCI", "KLEVE" "THE_QUICK_BROWN_FOX_JUMPS_OVER_THE_LAZY_DOG" and a unique sentence for each participant. The average values of the lines can be found above them.

was 8.74 cpm for unique sentences (see Fig. 2), while the n-gram model study showed an average of 18.9 cpm for their unique sentence tasks. This indicates that ChatGPT did perform worse compared to n-gram prediction model, but these differences may also be described by the ChatGPT communication delays.Future steps could involve the implementation of another API that is specifically designed for auto-correction or integrating a locally executed artificial-intelligence(AI) model tailored for automatic text correction, potentially addressing many of the data protection concerns previously mentioned, and also increasing the performance. Additionally, exploring the power consumption and comparing it to the n-gram model.

CONCLUSION

We successfully developed a cVEP based spelling interface that incorporates the ChatGPT API to assist users with spelling tasks. The integration of ChatGPT expands the software's functionality, potentially improving communication efficiency. However, it's noteworthy that its assistance didn't come close to that of an n-gram model in terms of output characters per minute (OCM) in unique sentence tasks. Despite this limitation, it simplifies future development and reduces processing power requirements for local machines, potentially enhancing the typing experience. Our study underscores the potential of BCI-AI collaboration to enhance communication, autonomy, and quality of life for individuals with physical disabilities, though further research is needed.

ACKNOWLEDGEMENT

The authors thank the financial support provided by "The Friends of the University Rhine-Waal - Campus Cleve" association. Additionally, we appreciate the student assistants and all participants in this study.

REFERENCES

[1] Wolpaw J, Wolpaw EW. Brain-Computer Interfaces: Principles and Practice. OUP USA (2012).
[2] Volosyak I, Rezeika A, Benda M, Gembler F, Stawicki P. Towards solving of the Illiteracy phenomenon for VEP-based brain-computer interfaces. Biomedical Physics & Engineering Express. 2020;6(3):035034.
[3] Gembler F, Volosyak I. A Novel Dictionary-Driven Mental Spelling Application Based on Code-Modulated Visual Evoked Potentials. Computers. 2019;8(2):33.
[4] Lappalainen Y, Narayanan N. Aisha: A Custom AI Library Chatbot Using the ChatGPT API. Journal of Web Librarianship. 2023;17(3):37–58.
[5] Sharma A, Bhargava M, Khanna AV. AI-Farm: A crop recommendation system. In: 2021 International Conference on Advances in Computing and Communications (ICACC). IEEE, Oct. 2021.
[6] Dave T, Athaluri SA, Singh S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in Artificial Intelligence. 2023;6.
[7] OpenAI. "OpenAI Documentation." (2022). [Online]. Available: `https://platform.openai.com/docs/overview` (visited on 02/21/2024).
[8] Gembler F, Stawicki P, Saboor A, Volosyak I. Dynamic time window mechanism for time synchronous VEP-based BCIs-Performance evaluation with a dictionary-supported BCI speller employing SSVEP and c-VEP. PLOS ONE. 2019;14(6):e0218177.

Published by
Verlag der Technischen Universität Graz