Wearable Brain-Computer Interface Paired with Large Language Model for Fast Speech Communication

Christopher Samra
University of Waterloo
csamra@uwaterloo.ca

Livia Murray
University of Waterloo
l8murray@uwaterloo.ca

Dhrumil Patel
University of Waterloo
dhrumil.patel@uwaterloo.ca

***Abstract*—Current non-invasive wearable brain-computer interfaces for communication are intended for people with neurological conditions who cannot otherwise communicate. One of the largest challenges with these existing solutions is that their information transfer rates are much lower than verbal speech. This limits the quality of life of those who use these interfaces as it prevents them from speaking as quickly as they would like to, and in many cases conversations and social interactions that are time sensitive unfairly exclude these users from participation. To minimize this problem, our team built an interface that pairs existing techniques in literature (Electroencephalography, Steady State Visually Evoked Potentials), with new advancements in Machine Learning (Large Language Models) to provide intelligent phrase suggestions based on conversational context. Preliminary results show our implementation has enabled people using our interface to communicate over two times faster than anything we could find in existing literature.**

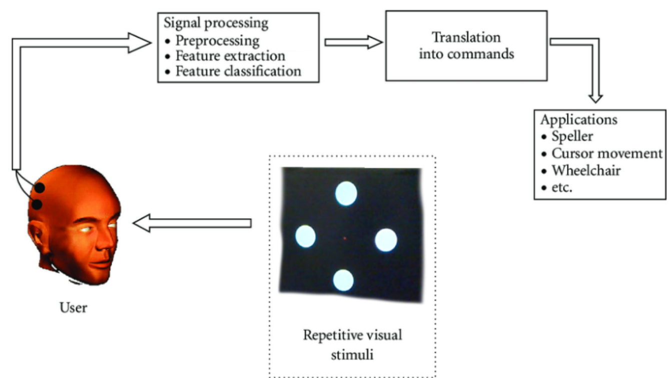

Fig. 1. SSVEP Interface Diagram

I. Introduction

A. Motivation

Brain computer interface (BCI) technology can provide a new channel of communication to humans, especially those with severe disabilities, such as stroke, spinal cord injury and amyotrophic lateral sclerosis (ALS) [1].

Many wearable BCI systems use electroencephalography (EEG) to measure brain activity. EEG measures brain activity non-invasively by recording electrical activity at the scalp that arises from large activations of brain cells [2]. A common paradigm used in EEG BCI literature is known as Steady State Visually-Evoked Potentials (SSVEP). SSVEP is a type of brain activity that occurs in response to a continuous visual stimulus flickering at a specific frequency [3]. SSVEPs are used as input signals to allow users to control elements of a user interface on a screen [3]. The basic principle behind SSVEP is that when a person focuses their attention on a visual stimulus that flickers at a specific frequency, their visual cortex produces a corresponding electrical signal at the same frequency [3].

One of the common applications for SSVEP interfaces include keyboards, more commonly known as SSVEP spellers [3].

B. Related Works

Traditionally, SSVEP speller designs consist of targets of various alphanumeric characters that are each assigned different frequencies of stimuli. These characters are selected when the user focuses on the stimuli associated with them [3]. However, while this may be acceptable for texting, typing character by character may not be ideal when BCI users want to speak to others verbally.

One implementation of an SSVEP speller attempted to use autocomplete to rectify this problem, but in their testing the researchers did not see information transfer rate (ITR) scores higher than 78 bits/min [4], this translates to around 5 words per minute [5].

The best case example of an SSVEP interface using characters to type that we could find yielded an average ITR of 151 bits/min [6], this translates roughly to 10 words per minute [5].

These ITRs are too slow to catch up with verbal communication rates, especially considering the average conversational verbal speech rate is between 120-150 words per minute [7].

C. Problem Definition

Almost all existing SSVEP speller applications that our team reviewed have proven to be not suitable for verbal communication due to their design restricting ITR.

SSVEP spellers have design limitations due to physical screen space, and the range of most usable stimuli frequencies (8-18hz) [8]. Because of these restrictions, only a limited number of interactable keys (buttons with stimuli) can be present at a given time, meaning character by character typing can be time consuming. This low ITR problem has a huge impact when people with neurological conditions attempt to communicate in time-sensitive social situations, and they can often feel left behind. In the next section we discuss the methodology we employ to build out our own SSVEP interface and how we use a Large Language Model (LLM) to improve ITR.

II. Methodology

A. BCI Design

For our EEG device, our team used a gtec Unicorn [9], we designed custom mounts for the electrodes to fit inside a modified OpenBCI UltraCortex Mk4 headframe [10], these design decisions allowed us to take advantage of the gtec's better signal clarity in a head-frame that could be fitted to several different head sizes.

The eight EEG channels provided by the device were placed on the PO4, O2, OZ, POZ, PO3, O1, PO8, PO7 EEG Electrode locations [11], chosen for proximity to the occipital region (the region responsible for processing visual stimulus).

Eight stimuli were chosen with the following frequencies 11.75, 12.75, 10.25, 14.75, 11.25, 8.25, 10.75 and 13.25 hz. These stimuli were chosen arbitrarily within the 8 to 15Hz range as this range yielded the best results in our testing.

Fig. 2. Subject Wearing EEG Headset

B. Data Collection and Validation

SSVEP data was collected and visualized offline to validate signal quality. This was done by applying a bandpass filter from 6 to 18 hz and visualizing the signals in two ways.

The first being a plot of the Fast Fourier Transform (FFT) of the EEG signal averaged across trials from a particular stimuli frequency and EEG channel, the second validation method involved taking the first Principal Component Analysis (PCA) component from one trial of a stimuli frequency across all channels. The response to the stimuli in the SSVEP data collection method was validated as being correct with a comparison between these two approaches.

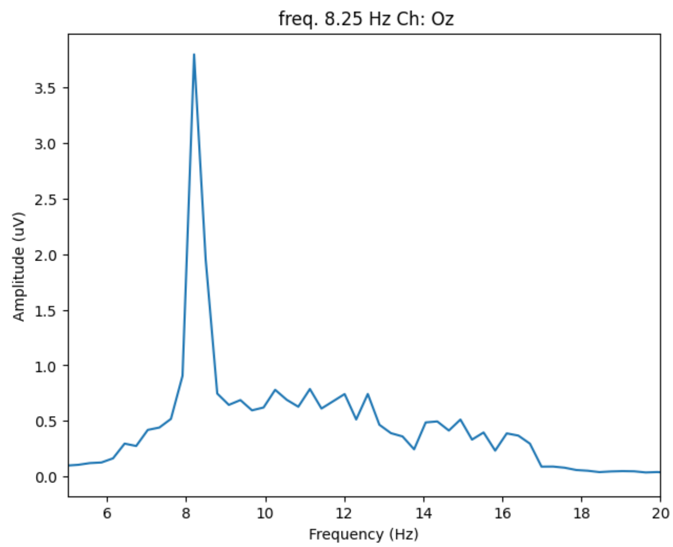

Fig. 3. FFT Averaged Across Trials

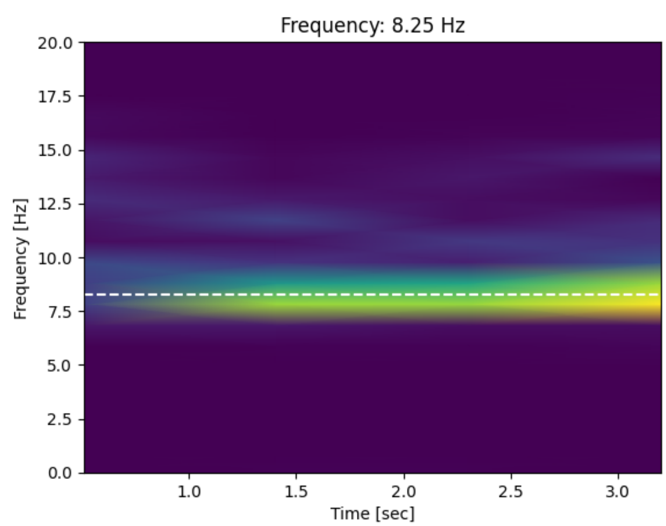

Fig. 4. PCA Across EEG Channels

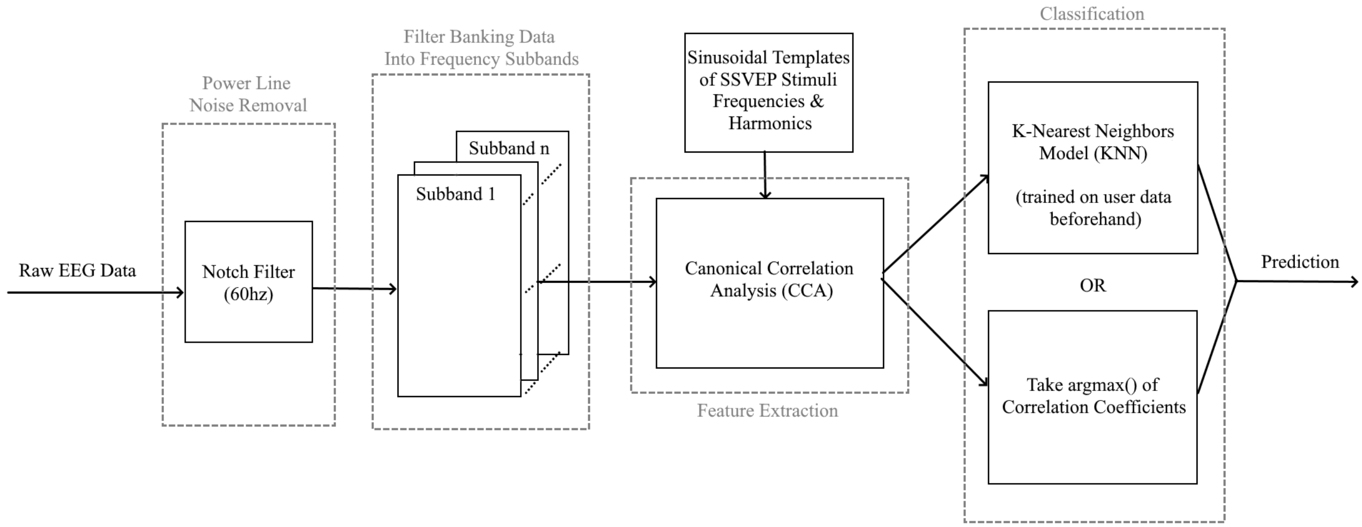

Fig. 5. EEG SSVEP Pipeline Diagram

C. EEG Processing and SSVEP Classification Pipeline

The pipeline ingests raw EEG data at 250 hz. A notch filter is applied at 60 hz to remove power line noise [12].

For further preprocessing and feature extraction we use Filterbank Canonical Correlation Analysis (FBCCA). This consists of two steps, with the first step being a filterbank technique that decomposes the signal into 10 frequency subbands following the "M3" subband decomposition from [6].

The second step involves using Canonical Correlation Analysis (CCA). CCA takes in both the processed EEG subbands and the sinusoidal reference templates matching the stimuli base and harmonic frequencies to identify correlation between the EEG signal and the stimuli frequencies [13].

To do this CCA computes pairs of coefficient vectors that correspond to projections on the EEG data and the reference templates [13]. We then compute a correlation statistic between these vectors using Pearson Correlation and create a matrix of correlation statistics between each frequency subband and the reference templates. We weigh the correlation statistics using a dot product with pre-defined weights that weigh lower frequency bands more heavily than higher frequency bands. This weighted correlation statistic matrix contains our features.

For classification, the system can either feed the features into a trained K-Nearest Neighbors (KNN) model (number of neighbors set to four) for stimuli prediction, or simply predict stimuli based on the max correlation statistic between a set of subbands and a reference template. For the rest of the paper, these approaches will be referred to as FBCCAKNN and FBCCA-max respectively.

D. Large Language Model Usage in SSVEP Interface

The figure below Illustrates the most common user flow of our application.

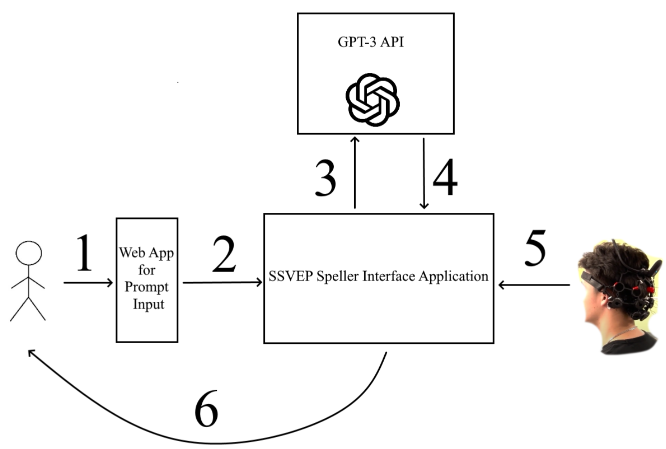

Fig. 6. SSVEP Interface User Flow Diagram

1) Able-bodied user talks to BCI user using speech to text or simply typing in a companion web app.
2) Message is sent to the interface.
3) Interface queries GPT-3 for phrase responses based on input message from 1.
4) GPT-3 returns phrase suggestions.
5) BCI user forms response with phrase suggestions.
6) Text to speech layer reads response aloud to able-bodied user.

For our SSVEP speller, instead of defaulting to characters, our keys defaulted to phrase suggestions obtained using OpenAI's GPT-3 API, more specifically, the davinci model was used for our tasks [14].

The entire SSVEP interface application, from data streaming, to the pipeline, to the graphical user interface was written in Python. This application and its web app companion can be found on our GitHub repository [15].

III. Results (Preliminary)

The first set of results were found by having two subjects navigate and type messages in our SSVEP interface in two separate sessions, one session using the FBCCAKNN classifier and one using the FBCCA-max classifier. It is important to note that for each subject, the KNN in FBCCAKNN was trained separately on the subjects data alone. The number of total key selections and correct key selections were recorded from each session, and classification accuracy metrics for each session were determined from this.

TABLE I
SSVEP Classification Performance

Subject	Classifier	Selections	Correct Selections	Accuracy
1	FBCCA-max	24	21	87.50%
1	FBCCAKNN	27	23	85.18%
2	FBCCA-max	36	30	83.33%
2	FBCCAKNN	31	28	90.32%

Based on the results from Table 1, it is unclear whether the FBCCAKNN classifier performed better than the unsupervised FBCCA-max approach, more data from more subjects would be needed to prove a significant difference in classification accuracy. Considering FBCCAKNN requires additional data collection and training beforehand, it is safe to say (at least with these early results) that FBCCA-max has a better setup effort to accuracy ratio compared with FBCCAKNN.

The next set of results come from a conversation we had with one subject while they used our SSVEP interface. We spoke back and forth with the subject three times, and they responded to each of our prompts using the phrase suggestions on the SSVEP interface that came from GPT-3.

TABLE II
ITR Performance from Example Conversation using GPT-3 responses (FBCCA-max)

Response	Time to respond (s)	Words	ITR (words/min)
1	30	17	34.00
2	35	10	17.10
3	62	15	14.50
AVG	42.67	14	21.87

Based on the results in Table 2, although preliminary, it is still very significant that on average in this example our system's ITR is over two times that of the highest we could find (10 words a minute [6]). This is very promising for the future of wearable speech BCIs, and demonstrates that Large Language Models can significantly improve ITR for these systems.

IV. Conclusion

To conclude, our team has shown that existing methods in EEG BCI literature along with the incorporation of LLMs work well together, and in some contexts can dramatically improve rate of speech communication. This ultimately brings us one step closer to minimizing the gap between a wearable BCI user's rate of communication and that of an able-bodied individual.

One of the disadvantages of our implementation however was that it did not have any personalization. This meant that it only took into account the prompts it was receiving in making phrase suggestions and it did not take into account environmental, personal, or relationship contexts when forming phrase suggestions, meaning more personalized messages would still take a lot longer to be communicated.

Although conversational context was taken into account, future research should adopt a speech language pathology perspective to work towards quicker and more personalized communication. This could be achieved by implementing an additional context engine, considering environmental and personal context, for example, the relationship the user has with the person they are communicating with.

References

[1] M. T. Medina-Juliá, Á. Fernández-Rodríguez, F. Velasco-Álvarez, and R. Ron-Angevin, "P300-based brain-computer interface speller: Usability evaluation of three speller sizes by severely motor-disabled patients," Frontiers in Human Neuroscience, vol. 14, 2020.

[2] "Electroencephalogram (EEG)," Electroencephalogram (EEG) — Johns Hopkins Medicine, 08-Aug-2021. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg. [Accessed: 12-Mar-2023].

[3] M. Li, D. He, C. Li, and S. Qi, "Brain–Computer Interface speller based on steady-state visual evoked potential: A review focusing on the stimulus paradigm and performance," Brain Sciences, vol. 11, no. 4, p. 450, 2021.

[4] N. Shi, L. Wang, Y. Chen, X. Yan, C. Yang, Y. Wang, and X. Gao, "Steady-state visual evoked potential (ssvep)-based brain–computer interface (BCI) of Chinese speller for a patient with amyotrophic lateral sclerosis: A case report," Journal of Neurorestoratology, vol. 8, no. 1, pp. 40–52, 2020.

[5] "Units Converters," unitsconverters.com, 2016. [Online]. Available: https://www.unitsconverters.com/en/Bitpersecond-To-Wordperminute/Unittounit-4798-7602. [Accessed: 12-Mar-2023].

[6] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter Bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," Journal of Neural Engineering, vol. 12, no. 4, p. 046008, 2015.

[7] D. Barnard, "Average speaking rate and words per minute," VirtualSpeech, 08-Nov-2022. [Online]. Available: https://virtualspeech.com/blog/average-speaking-rate-words-per-minute. [Accessed: 12-Mar-2023].

[8] R. Kuś, A. Duszyk, P. Milanowski, M. Łabecki, M. Bierzyńska, Z. Radzikowska, M. Michalska, J. Żygierewicz, P. Suffczyński, and J. Durka, "On the Quantification of SSVEP Frequency Responses in Human EEG in Realistic BCI Conditions," PloS one, vol. 8, no. 10, Oct. 2013.

[9] "Wireless 8-channel EEG headset," Unicorn Hybrid Black, 12-Jan-2023. [Online]. Available: https://www.unicorn-bi.com/. [Accessed: 12-Mar-2023].

[10] "Ultracortex 'Mark IV' EEG headset," OpenBCI Online Store. [Online]. Available: https://shop.openbci.com/products/ultracortex-mark-iv. [Accessed: 12-Mar-2023].

[11] F. Lotte, L. Bougrain, and M. Clerc, "Electroencephalography (EEG)-based brain-computer interfaces," Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1–20, 2015.

[12] S. O. Gilani, Y. Ilyas, and M. Jamil, "Power line noise removal from ECG signal using notch, band stop and adaptive filters," 2018 International Conference on Electronics, Information, and Communication (ICEIC), 2018.

[13] M. Nakanishi, Y. Wang, Y.-T. Wang, and T.-P. Jung, "A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials," PLOS ONE. [Online]. Available: https://doi.org/10.1371

[14] OpenAI API. [Online]. Available: https://platform.openai.com/docs/models/gpt-3. [Accessed: 12-Mar-2023].
[15] "Watolink," GitHub. [Online]. Available: https://github.com/WATOLINK. [Accessed: 12-Mar-2023].