Mitchell Bartell

NEUR 4001

Individual Capstone Project

Objective

The mission of this paper is to explore the future of neurotechnology by examining EEG-based decoding of imagined speech, specifically the mental articulation of letters and short words. The goal is to develop a comfortable, non-invasive headset that leverages real-time ML models to decode imagined speech and restore communication for individuals with paralysis or speech loss. The headset would integrate EEG signal acquisition, real-time AI decoding, and speech synthesis to create an intuitive communication interface that redefines assistive technology.

Introduction

The field of cognitive speech extraction is challenged by physiological artifacts distorting electroencephalogram (EEG) signals (Ganushchak et al., 2011). These distortions limit the reproducibility and accuracy of phonetic decoding models across subjects, with some studies reporting accuracies as low as 41% (Lee et al., 2023). Metzger et al. (2023) demonstrated decoding attempted speech using a 256-channel ECoG system, achieving 62 words per minute with a 25% error rate. This shows that subdural, non-penetrating electrode arrays targeting the ventral M1 and IFG regions can achieve conversational speeds. Building on this, a 2024 UCSF/Stanford collaboration introduced a "brain-to-voice" decoder that reconstructs a patient's natural voice in real time using pre-injury vocal profiles. (Défossez et al., 2023) extended these efforts using self-supervised learning to decode heard speech from EEG signals in 175 participants. Despite obstacles, recent developments in machine learning (ML) models such as ChatGPT offer a promising avenue for filtering out physiological noise and extracting reliable event-related potentials (ERPs) related to phonetic cognition. EEG is a non-invasive brain-computer interface (BCI) modality that detects cortical frequency band activity when electrodes are placed over relevant brain regions. Machine learning plays a pivotal role in analyzing EEG-based BCI data to detect patterns of neural activation associated with specific cognitive tasks (Cao, 2021). Across studies, the common challenge remains: decoding accuracy and fidelity are bottlenecks in real-time phonetic communication via neural interfaces.

Hypothesis

We hypothesize that phonetic cognition elicits distinct, reproducible event-related potentials (ERP) waveforms in EEG recordings that can be reliably detected and decoded by an ML model to transduce imagined speech.

Method

This experiment consists of individual subject data collection, data processing, and ML training. Fitted with a 32-Channel FLEX 2 Gel EEG headset, a participant will be shown letters in the English alphabet and be instructed to think about that letter. This could be spelling out the word cognitively or thinking about the meaning of the word. This will be Task 1 and will last 3 minutes. A 30-second delay is implemented to reset cognitive activity. For Task 2, the participant will be shown four simple 3-letter words, such as “CAT” and be instructed to think about constructing the word metaphysically. Task 2 will last 3 minutes. Then, another 30-second delay palate cleaner. For Task 3, the participant will be instructed to focus and think on just one of four words. After the experiment, the data will be pre-processed. The data set will be loaded into EEGLAB in MATLAB. The event file will be imported to check for event markers and to filter out noise or artifacts. The event time range in EEGLAB will be 200. Noise and artifacts will be filtered with a cut-off frequency of 1 Hz and 50 Hz. High-pass motion artifacts below 1 Hz will be removed, and motion artifacts filtered. Low-pass line noise above 50 Hz will be removed, and line noise artifacts filtered. Furthermore, each channel except 32-channels will be removed to extract only related data.

Mitchell Bartell

NEUR 4001

Individual Capstone Project

The legacy reasoning model o1 pro of ChatGPT will be used to interpret the data, looking for non-overlapping reproducible signals during the first task. Specifically, the model will be instructed to comb the dataset for phonetic cognition ERP waveforms. The engine will be prompted with "Use this data set of a subject of an EEG cognitive phonetic production experiment. Comb through the data and extract ERP waveforms of the subject during the task. Specifically, use the given timestamps to record non-overlapping and reproducible EEG signals. Once the model is trained, the ML model will then be used to analyze the periods in which the participant experienced phonetic cognition during task 2. The model will be prompted to "With the signals found in Task 1, extract and record sequential non-overlapping, cross-referenced EEG signals given these timestamps".

In summary, the experimental design includes three structured tasks to elicit phonetic cognitive activity in participants:

1. **Task 1 (Phoneme Imagination)**: The participant will view single English letters and be instructed to think about the letter, spelling it out mentally or contemplating its meaning. Duration: 3 minutes. 30-second rest.

2. **Task 2 (Word Construction)**: The participants will view simple three-letter words (e.g., "CAT") and be instructed to mentally construct or imagine the word as a phonetic unit. Duration: 3 minutes. 30-second rest.

3. **Task 3 (Word Selection)**: The participants will be asked to think deeply about only one of the four previously seen words. Duration: 3 minutes.

After each task, EEG data will be preprocessed using EEGLAB in MATLAB:

- Import the dataset and check event markers using ML model
- Filter noise/artifacts using a 1 Hz high-pass and 50 Hz low-pass band
- Remove motion artifacts below 1 Hz and line noise above 50 Hz
- Retain only the relevant 32 channels

Following pre-processing, the dataset will be analyzed using a reasoning-capable language ML model (legacy ChatGPT O1 Pro). The model will be prompted to "Identify ERP waveforms during Task 1 that are temporally non-overlapping and reproducible. Apply the ERP waveform features from Task 1 to detect corresponding phonetic cognition during Task 2. Finally, isolate distinct patterns corresponding to specific imagined words during Task 3 and approximate the imagined word".

Predicted Results

The predicted result of this experiment is that ERP waveforms will be detected and linked to imagined speech in the form of letters, three-word letters, and form a framework that can deduce higher levels of speech. We predict that reproducible EEG signals corresponding to phonetic cognition will be detectable during all three tasks. Specifically, Task 1 will yield reliable ERP features for single-letter phoneme imagination. Task 2 will demonstrate the model's ability to link ERP patterns to the phonetic construction of simple words. Lastly, Task 3 will validate whether isolated word selection can be mapped to distinct EEG patterns. If the ML model correctly approximated the subject's word, a framework for imagined speech decoding with an accuracy rate 95%+ could be possible.

Mitchell Bartell

These results show that imagined speech can be transduced by an ML model and interfaced to a non-invasive BCI apparatus in real time. Real-time monitoring of imagined speech would allow for patients suffering from paralysis or the loss of speech a new way to communicate. This technology could be improved to be a cognitive speech translator, allowing people to translate their imagined speech to any foreign language. Interestingly, this technology could be extended to predict real-time behavior and have defense applications. The technology could allow warfighters to covertly communicate in stressful environments using their minds. Furthermore, the technology could monitor combat readiness of pilots, special warfare operators, and larger structural units such as divisions, significantly improving tactical efficiency in military operations. Real-time, non-invasive BCI technology, as described, would also allow warfighters to control tactical equipment using their minds.

Discussion

There is a possibility that no ERP waveforms are extracted during the experiment. The ML model could not be effective at filtering physiological artifacts. If this were the case, data processing would include recording and removing physiological EEG signals using the ML model before data analysis. The ML model could require additional training if no signals are found. Additional training would include repetitive identification tasks. Alternatively, a different, more specialized ML model could be more effective than OpenAI's ChatGPT. In this case, data would simply be imported to the new ML model and trained accordingly. If remediations are not sufficient, the results of the experiment would be due to random chance, and further adjustments would be required.

Mitchell Bartell

NEUR 4001

Individual Capstone Project

Limitations

This experiment is limited; imagined speech could be difficult for the ML model to detect. Narrow bands of signal could be misinterpreted as imagined speech signals and distort data. This could lead to a random word selection in Task 3 and misrepresented data in Task 1 and Task 2. Extensive backtesting, validation, and study expansion would diminish the chance of a random or misrepresented result. There is a possibility that EEG signals produced by imagined speech could be invisible or hidden in larger signal structures. In this case, higher resolution EEG technology would be necessary to extract imagined speech signals. Furthermore, an expansion of the subject size would lead to a more conclusive dataset, since imagined speech is more than likely unique to each individual.

Mitchell Bartell

NEUR 4001

Individual Capstone Project

References

Cao, Z. (2020). A review of Artificial Intelligence for eeg‐based brain−computer interfaces and applications. *Brain Science Advances*, *6*(3), 162–170. https://doi.org/10.26599/bsa.2020.9050017

Défossez, A., Caucheteux, C., Rapin, J. *et al.* Decoding speech perception from non-invasive brain recordings. *Nat Mach Intell* 5, 1097–1107 (2023). https://doi.org/10.1038/s42256-023-00714-5

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of Electroencephalography in language production research: A Review. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00208

Lee, Y.-E., Lee, S.-H., Kim, S.-H., & Lee, S.-W. (2023). Towards voice reconstruction from EEG during imagined speech. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(5), 6030–6038. https://doi.org/10.1609/aaai.v37i5.25745

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., Dougherty, M. E., Liu, J. R., Wu, P., Berger, M. A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G. K., & Chang, E. F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, *620*(7976), 1037–1046. https://doi.org/10.1038/s41586-023-06443-4