Imagined Speech BCI with EEG and Deep Learning

Qihao Xu

mc36420@umac.mo

CCBS

ICI

University of Macau

I. Introduction

Brain-computer interface (BCI) is a kind of technique providing a bridge between the human brain and computer to convey information. There have been many studies investigating the application of BCI in many areas like healthcare, entertainment, gaming, communication, control, research and so on. As the concept of connected environments being popular, BCI has been one of the frontiers of research interest. The implementation of BCI nowadays usually uses deep learning techniques to analyze the brain signals from devices like ECoG, MEG, EEG, fMRI, fNIRS and so on to decode brain commands to control computers, prosthetic arms and so on. Compared with other devices, EEG is safe, cheap, portable and reliable enough which makes it a good choice for BCI. DL (Deep learning) methods are popular at this time for high accuracy and good generalization compared with conventional machine learning algorithms. With the classification method of deep learning combined with other steps of BCI data processing: signal acquisition, pre-processing, feature extraction, a BCI can be implemented.

Imagined speech BCI is a newer kind of BCI compared with motor imagery and other BCI. It receives brain signals from brain regions related to speech and language and decoding brain signals into voice or text for convenient control of devices or making people who cannot speak communicate with others.

Though transmitting information from computer to brain still seems impossible for now, transmitting signals from the human brain to computer is already practical enough. With this technique, we can give paralyzed people one more chance to speak and control devices. Also, normal people can also use imagined speech BCI to increase their efficiency of working because of the ability of convenient control and communication. We can also get deeper understanding of neuroscience as the development of brain signal analysis.

The objective of this review is to delve into the recent advancements in utilizing DL techniques for speech imagery decoding using EEG signals, highlighting emerging trends, challenges, and prospects. Our major contributions are as follows:

⚫ Detailed publicly accessible SI EEG datasets along with experimental paradigms, pre-processing approaches, and input data formulation methods for DL.

⚫ Provided practical guidelines for developing efficient DL models tailored to SI decoding by optimal consideration of all the DL design trends.

⚫ Highlighted challenges in technical usability and application of DL-based SI decoding.

⚫ Explored future prospects of using EEG signals with DL for the effective decoding of SI.

⚫ Provided a detailed table that precisely scrutinizes speech imagery decoding literature.

The review is organized as follows. Section I introduces the background of this issue, expounds on the motivation of this review and distinguishes the characteristics of this article Section II lists the search strategy and inclusion criteria. Section III examines key components of the reviewed papers, including EEG datasets, preprocessing strategies, input formulations, DL architectures, and methods of performance evaluation. Section IV elaborates on major finding, highlights the theoretical and practical challenges of DL-based imagined speech decoding and suggests future research directions. Section V concludes and summarizes this review.

II. Methods

A. Search methods for identification of studies

The literatures are collected from reviews and papers in Google Scholar with the prompt of ‘Imagined Speech’, ‘BCI’, ‘Deep Learning’, and ‘EEG’. These search terms were sufficient to retrieve all relevant studies.

After retrieving the search results, we moved the duplicates between databases and then carefully identified the remaining articles based on the following criteria.

Articles inclusion criteria:

⚫ Written in English.

⚫ Deep learning models used for decoding imagined speech.

⚫ EEG signals within.

⚫ Tasks based on decoding imagined speech.

We retrieved 100 research articles in total.

B. Data extraction and presentation

This review focuses on the important information that counts in the field of decoding speech imagery using EEG signals and DL methods. The following categories were collected: dataset, feature extraction method, deep learning architecture, and performance.

III. Techniques of Imagined Speech Decoding

This section demonstrates the main findings from the reviewed studies, including dataset, feature extraction, DL architecture, and performance evaluation.

1. Dataset

The reviewed studies employed a total of 13 different open access imagined speech EEG datasets. This allowed researchers to compare their results with those reported by other authors and assess the reproducibility and generalizability of their findings. Table I illustrates the publicly available datasets utilized in the reviewed studies.

We surveyed the 13 public datasets that were used in the studies we reviewed. For all the datasets, the following characteristics were extracted: the number of subjects, language, the number of electrodes, types of stimuli, equipment, vocalization, types of prompts, and sampling rate.

We detail the 13 found public datasets:

Dasalla DB [1]: the dataset contains recordings from 3 subjects (aged 26 to 29, 2 male) who performed vowel imagery task (English vowels /a/ and /u/), with 64 EEG electrodes placed over the scalp at 10-20 system. The experimental paradigm begins with a beep to signal the start of a pre-stimulus period lasting 2 s. Then a visual cue of vowels was presented for 2-3 s, followed by a 3 s rest interval. 50 trials were performed for each task, resulting in a total of 150 trials per subject.

图示

描述已自动生成

Kara One [2]: the dataset combined 3 modalities (EEG, facial, and audio) from 12 healthy subjects (mean age 27.4, 8 male) during thinking and speaking of seven phonemes: /iy/, /uw/, /piy/, /diy/, /tiy/, /m/, /n/ and four words: “pot”, “pat”, “knew” and “gnaw”. 64-channel Neuroscan Quick-cap with electrode placed based on the 10-20 system was used to collect EEG data. All signals were collected using a standard protocol consisting of (1) a rest state, (2) a stimulus state, in which the prompt with the stimulus appeared, (3) a 5-second stimulus imagined speaking state, and (4) an actual articulation of the stimulus. In total, 132 trials were conducted, 12 for each prompt.

日程表

描述已自动生成

Coretto DB [3]: an open-access EEG signal database comprising data from 15 participants (mean age 25, 8males), obtained during the mental imaging of two groups: vowels (/a/, /e/, /i/, /o/, and /u/) and Spanish words (corresponding to up, down, right, left, forward, and backward, respectively). An acquisition system with six channels was used to record EEG signals at a sampling rate of 1024 Hz. During the presentation of target stimuli, four intervals were included: (1) A 2 s ready interval. (2) A 2 s stimulus presentation state, in which both visual and acoustic stimuli are given. (3) A 4 s imagine/pronounce stage, in which an image displays the task. (4) A 4 s rest interval. Each subject performs 50 trials for each word, 40 for imagined speech, and 10 for pronounced speech.

图示

描述已自动生成

ASU DB [4]: a new open-access EEG signal database comprising recordings from 15 subjects (aged from 22 to 32, 11 male). The subjects performed three different types of imagined speech: imagined speech of short words (“in”, “out”, and “up”), longwords (“cooperate” and “independent”), and vowels (/a/, /i/ and/u/). The EEG signals were acquired using a BrainProducts ActiCHamp amplifier with all 64 electrodes placed based on the10/20 international system. A beep sound appeared first when the trial started and was repeated four more times. At the beginning of the trial, the subject was also prompted with a visual cue indicating the desired word to be imagined. The cue lasted for 7 × T s. The subject was instructed to perform speech imagery at each beeping sound and continue at the same rhythm until the visual cue disappeared. Finally, the trial ended with a rest period of 2 s where no cue and no sounds were present. For short words and vowels, the period was T = 1 s, while for long words the period was T = 1.4 s. Each prompt was attempted 100 times by each subject, resulting in a total of 1200 trials for long words with six subjects (S2, S3, S6,S7, S9, and S11), 1800 trials for short words with six subjects (S1,S3, S5, S6, S8, and S12), 1120 trials for short-long words with six subjects (S1, S5, S8, S9, S10, and S14, the last two subjects performed 80 trials), and 2400 trials for vowels with eight subjects (S4, S5, S8, S9, S11, S12, S13, and S15).

图片包含 日程表

描述已自动生成

Nieto DB [5]: the dataset contains ten healthy right-handed participants, four females and six males with mean age = 34 (std = 10 years), with no hearing loss, no speech loss, and with no neurological, movement, or psychiatric disorders, joined the experiment and gave their written informed consent. All participants were native Spanish speakers. Within each session, five stimulation runs were presented. Those runs correspond to the different proposed conditions: pronounced speech, inner speech and visualized condition. At the beginning of each run, the condition was announced in the computer screen for a period of 3 seconds. In all cases, the order of the runs was: one pronounced speech, two inner speech and two visualized conditions. A one-minute break between runs was given (inter-run break). The classes were specifically selected considering a natural BCI control application with the Spanish words:“arriba”, “abajo”, “derecha”, “izquierda” (i.e. “up”, “down”, “right”, “left”, respectively). The trial’s class (word) was randomly presented. Each participant had 200 trials in both the first and the second sessions. Each trial began at time t = 0 s with a concentration interval of 0.5 s. The participant had been informed that a new visual cue would soon be presented. A white circle appeared in the middle of the screen and the participant had been instructed to fix his/her gaze on it and not to blink, until it disappeared at the end of the trial. At time t = 0.5 s the cue interval started. A white triangle pointing to either right, left, up or down was presented. The pointing direction of the cue corresponded to each class. After 0.5 s, i.e. at t = 1 s, the triangle disappeared from the screen, moment at which the action interval started. The participants were instructed to start performing the indicated task right after the visual cues disappeared and the screen showed only the white circle. After 2.5 s of action interval, i.e. at t = 3.5 s, the white circle turned blue, and the relax interval began. The participant had been previously instructed to stop performing the activity at this moment, but not to blink until the blue circle disappears. At t = 4.5 s the blue circle vanished, meaning that the trial has ended. A rest interval, with a variable duration of between 1.5 s and 2 s, was given between trials. To evaluate each participant’s attention, a concentration control was randomly added to the inner speech and the visualized condition runs. The control task consisted of asking the participant, after some randomly selected trials, which was the direction of the last class shown. The participant had to select the direction using the keyboard arrows. Electroencephalography (EEG), Electrooculography (EOG) and Electromyography(EMG) data were acquired using a BioSemi ActiveTwo high resolution biopotential measuring system (https://www.biosemi.com/products.htm). For data acquisition, 128 active EEG channels and 8 external active EOG/EMG channels with a 24 bits resolution and a sampling rate of 1024 Hz were used.

日程表

描述已自动生成

BCI Competition V-3 [6]: the dataset contains recordings from 15 subjects who performed word imagery tasks (English words “hello”, “help me”, “stop”, “thank you”, and “yes”). The EEG recordings were obtained using a 64-electrode cap positioned over the scalp according to the 10-20 system. The experimental paradigm used in the dataset is that the auditory cue for five words was presented randomly for 2 s. Each cue was followed by four consecutive rounds of a 1 s cross- mark interval and a 2 s imagined speech interval. Then, a 3s relaxation was given to clear the mind. 70 trials were performed for each word, resulting in a total of 350 trials in the dataset.

图示

低可信度描述已自动生成

FEIS [7]: The FEIS (Fourteen-channel EEG for Imagined Speech) dataset, comprises EEG recordings of 21 English-speaking participants recorded with a lightweight, 14-channel mobile headset with dry electrodes (the Emotiv EPOC+) at a sampling frequency of 256. Recordings are time-aligned with phone stimuli, consisting of three stimulus types: heard, spoken internally (imagined) and spoken overtly. Data collection methodology is adapted from that of the Kara One dataset [2]. Sixteen English phonemes were chosen to represent a balanced categorical spread of binary phonological features ([±nasal], [±back], [±voice], etc). They are shown in the tables below.

表格

描述已自动生成

A single instance of each of the 16 phones was recorded at 44.1 kHz with a cardioid microphone. The Emotiv EPOC+ is a mobile headset with semi-flexible sensor “arms” which allow for universal fitting, within a fixed configuration. While this allows ease of use, it means that electrode positions are inconsistent relative to the international10-20 montage system, due to participants’ different head sizes. The EEG recordings consist of 160 trials, comprising 6phonemes × 10 repetitions, randomized to maintain participant attention. Each trial has four 5-second “epochs”. First, a “resting” epoch, in which participants are shown the word “REST” on screen and attempt to clear their mind (resting state measurement can be used for task-specific feature extraction and reduces cognitive load). Next, a“stimuli” epoch, in which participants are played their own vocalization of a single phone looped five times, and shown a corresponding IPA representation (which participants were familiar with). Next, a “thinking” epoch, in which participants are presented with a blank screen, and imagine repeating the phone, but without any articulator movement. Finally, a “speaking” epoch, in which participants are prompted with an image of a mouth to then vocalize the phone. In each of the two latter epochs, subjects imagine/speak the phone five times in a steady rhythm, imitating the recording played in the stimuli epoch.

图形用户界面, 图示

中度可信度描述已自动生成

Simistira DB [8]: This is a publicly available bimodal dataset containing EEG and fMRI data acquired simultaneously during inner-speech production. Data was obtained from four healthy, right-handed participants during an inner-speech task with words in either a social or numerical category. Each of the 8-word stimuli were assessed with 40 trials, resulting in 320 trials in each modality for each participant. The EEG data were acquired using the BioSemi Active2 measuring system (BioSemi B.V., Amsterdam, Netherlands) with a 16-bit resolution and a sampling rate of 512 Hz. A BioSemi EEG head cap with 64 electrodes in pre-fixed electrode positions and 6 external sensors were used. Two categories, social and number, with four words each were selected for imagined speech paradigm. The social category contained the words child, daughter, father, and wife. The number category contained the words four, three, ten, and six. The EEG recordings consisted of one session with 40 trials per word. A fixation period of 2 s was followed, in which the participants were instructed to fixate their eyes on the center of the screen. The 2 s fixation time was included only once at the beginning of each session since there was a significant rest period of 10 seconds between trials. Then, each trial consisted of the inner-speech task (4 s) and a subsequent rest period (10 s). Eight different words were used for the inner-speech task, divided into 2 categories (social or number words), and there were 20 trials for each word in each session; thus, each session consisted of 160 trials. During the inner-speech task, the word stimulus was presented in white font against a black background for 4 s, and the participants were encouraged to repeat the given word in their minds as many times as possible (approximately 4 times) without any accompanying articulation or muscle movement (i.e., using their inner speech). The word stimuli were presented in a random order over the 160 trials. During the rest period, a white fixation cross was presented for 10 s, and the participants were allowed to relax and prepare for the next trial. The total duration of the recordings for the 320 repetitions was 74.6 min per participant.图形用户界面, 应用程序

描述已自动生成

DAIS [9]: DAIS database is an open database consisting of electroencephalography(EEG) and speech data from 20 participants recorded during the covert (imagined) and actual articulation of 15 Dutch prompts. Twenty native Dutch speakers, 14 women and 6 men (mean age: 24.6 years, SD: 1.0, range 23-26) participated. The prompts consisted of five Dutch vowels (/a:, e:, o:, i, u/, where “:” indicates lengthening) and ten Dutch words. The ten words are five Dutch word-pairs that are also words when read backwards: taal, laat, leeg, geel, niet, tien, toon, noot, soep, poes(Eng: “language”, “late”, “empty”, yellow”, “not”, “ten”, “tone”,“note”, “soup”, and “cat”). During the entire experiment, continuous EEG was collected from 64 electrodes using the TMSi SAGA 64+ (with a BrainWave EEG Cap using the 10-20 system) at a sampling frequency of 1024 Hz. The articulated speech was recorded using an Audio Technica AT2020USB+ microphone (F s= 44.1 kHz). Each experiment consisted of 20 runs of 15 trials, one for every prompt (i.e., the 15 Dutch vowels and words). The prompts were randomized for each participant using a balanced Latin square to reduce order effects. A trial consisted of four successive segments: pre-stimulus (rest; blank screen), reading of the prompt (either a vowel, denoted in orthographic script for ease of the participant, or a word), covert (imagined) speech (indicated with a thought balloon), and articulated speech (indicated with a speech balloon). Each run was followed by (another) 2s rest.

图示

描述已自动生成

LaRocco DB [10]: This is an open access imagined speech dataset with complete phoneme. A total of 16 participants were recruited through word of mouth and flyers. The average age was 27.3 ± 1.2 years, and the participants comprised 4 females and 12 males. Participants had normal hearing and normal or corrected vision. Most participants were native English speakers, with 5 non-native English speakers who demonstrated functional proficiency on the standardized exams required for university admission. Data recording for each phoneme included a demonstration and five separate trials. The phoneme was shown for 2 s. An auditory pronunciation of the phoneme was played once, at the start of each phoneme. A screen with the word‘wait’ was presented as an interval for 1 s. The phoneme was displayed again for 2 s, and the participant was instructed to think of speaking during this time. The character representation for each phoneme was then put on screen for 2 s, followed by a ‘wait’ screen for 1 s. Participants were instructed to stop imagining during the‘wait’ interlude screen, and the corresponding EEG data from these interlude segments was not used. This sequence was repeated five times without auditory feedback for five trials per phoneme. Each session included five trials of 44 phonemes in random order. To facilitate integration with open–source hardware and software, data were acquired using an OpenBCI Cyton board and an Ultracortex Mark IV headset (OpenBCI Foundation, New York). Data from 16EEG channels were acquired at 250 Hz. The10–20 International System electrode channels used included Fp1, Fp2, F7, F3, F4, F8, T3, C3, C4, T4, T5, T6, P3, P4, O1, and O2

图示

描述已自动生成

Sarmiento DB [11]: This is an open access dataset with 50 individuals, recorded under controlled conditions, for imagined vowels (/a/,/e/,/i/,/o/,/u/). This new database held the information of 50 university students (20 women and 30 men) whose native language is Spanish (M = 24.76, SD = 7.66). The signals were recorded with a 14-channel EMOTIV EPOC+ amplifier, with a sampling frequency of 128 Hz, a 14 bits resolution with 1 LSB with 0.51 μV in monopolar configuration, the original name of the Emotiv electrodes is as follows: E1 (AF3), E2 (F7), E3 (F3), E4 (FC5), E5 (T7), E6 (P7), E7 (O1), E8 (O2), E9 (P8), E10 (T8), E11 (FC6), E12 (F4), E13 (F8), E14 (AF4). For the experiment, each subject was told to imagine a given vowel continuously without pronouncing it while the light source was on. They were also told that, when the light source was turned off, they had to stop imagining the vowel and relax their body. During the experiment, the light source remained on for four seconds and then was turned off for three seconds. The procedure was repeated 25 times for each one of the imagined vowels. Upon completion of the 25 imagined speech tasks for each vowel, subjects rested for 5 min to continue with the next vowel. The imagined tasks were arranged in the following order: /a/,/e/,/i/,/o/,/u/.

图表

中度可信度描述已自动生成

Jahangiri DB [12]: In the dataset, there are 10 neurologically healthy volunteers in the age group of 21–33. Four phonetically dissimilar phonemic structures “BA”, “FO”, “LE”, and “RY” were chosen as covert speech tasks. The experiment consists of 4 recording runs for a participant, each containing 30 trials of only one class. For all classes, an identical “beep” sound was used as the auditory cue. The user was informed of the task before each run and asked to covertly speak when they heard the timing cue. A random rest period between 3 and 7 s was placed between trials to prevent the user from anticipating onset time based on rhythm. The EEG signals were recorded using a 64 channel Biosemi ActiveTwo™ system. Electrode placement was done per the international ABC system, which for 64 channels corresponds to the10/10 system. The data were recorded at a sampling rate of 2048 samples/s, with guaranteed data frequency content of 0-409 Hz according to BioSemi.

图示

描述已自动生成

Kumar DB [13]: In this dataset, 23 participants aged between 15 to 40 years have been enrolled to collect data. EEG recordings have been performed using Emotiv EPOC+ sensor. A slide presentation was prepared that consisted of 20 text and 10 non-text items. Out of 20 text classes, 10 slides are of digits from 0–9, whereas the rest of the slides consist of 10 character images. A pictorial representation of 20 text classes is depicted below. From the dataset, each slide has been shown to every participant for 10 s. Next, the participant has been asked to envision the shown item for 10 s in eyes closed resting state. Between two successive recordings, a gap of 20 s has been introduced to clear the previous imaginary thoughts of participant. Using this protocol, 690 (i.e. 23×30) EEG recordings have been collected.

图片包含 白板

描述已自动生成

TABLE I

Open Access EEG Datasets used for Imagined Speech

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Subject  Number | | Vocalization | Prompt | Language | Stimulus | Equipment | Channels | Sampling Frequency |
| DaSalla DB [1] | | **3** | **Covert** | **Vowels** | **English** | **Visual** | **BioSemi ActiveTwo** | **64** | **2048** |
| Kara One [2] | | **12** | **Overt and Covert** | **Syllables and Words** | **English** | **Visual and Auditory** | **Neuroscan Quick-cap** | **64** | **1000** |
| Coretto DB [3] | | **15** | **Overt and Covert** | **Vowels and Words** | **Spanish** | **Visual and Auditory** | **Grass R 8-18-36** | **18** | **1000** |
| ASU DB [4] | | **15** | **Covert** | **Vowels and Words** | **English** | **Visual** | **ActiCHamp** | **64** | **1000** |
| Nieto DB [5] | | **10** | **Overt and Covert** | **Words** | **Spanish** | **Visual** | **BioSemi ActiveTwo** | **128** | **1000** |
| BCI Competition V-3 [6] | | **15** | **Covert** | **Words and Phrases** | **English** | **Auditory** | **Emotiv EPOC+** | **14** | **256** |
| FEIS [7] | | **21** | **Overt and Covert** | **Vowels Syllables and Words** | **English and Chinese** | **Auditory** | **Emotiv EPOC+** | **14** | **256** |
| Simistira DB [8] | | **4** | **Covert** | **Words** | **English** | **Visual** | **BioSemi ActiveTwo** | **64** | **512** |
| DAIS [9] | | **20** | **Overt and Covert** | **Vowels and Words** | **Dutch** | **Visual** | **TMSi SAGA 64+** | **64** | **1000** |
| LaRocco DB [10] | | **16** | **Covert** | **Syllables** | **English** | **Visual and Auditory** | **OpenBCI Cyton and Ultracortex Mark IV** | **16** | **256** |
| Sarmiento DB [11] | | **50** | **Covert** | **Vowels** | **Spanish** | **Visual** | **Emotiv EPOC+** | **14** | **128** |
| Jahangiri DB [12] | | **10** | **Covert** | **Syllables** | **English** | **Auditory** | **BioSemi ActiveTwo** | **64** | **2000** |
| Kumar DB [13] | | **23** | **Covert** | **Words** | **English** | **Visual** | **Emotiv EPOC+** | **14** | **2000** |

B. Feature Extraction

Calculating features are considered as a traditional method for acquiring better performance in deep learning. The features are extracted from the raw EEG data and then put into deep learning models for classification. There are 31% of reviewed studies that don’t use feature extraction with only time-series raw data. Some of them use image as input like CWT (5%) and STFT(3%). The most used feature extraction method are CSP and DWT, with 11% of the reviewed studies and 10% of the reviewed studies. And there are also some less used methods like FFT (6%), Statistics features (6%), Entropy features (5%). And there are also other much less used feature extraction methods that take the proportion of 23%. We can see that about 1/3 of the studies don’t use feature extraction methods, and there isn’t any feature extraction method to be the main-stream method.

#Different methods in one study are counted separately

##Others means the method does not include any of the method in the figure

C. Deep Learning Architecture

In this section, we focus on DL architectures employed in EEG-based imagined speech decoding research. DL models are categorized according to their function into three subcategories: discriminative, representative, generative. Also, there are Machine Learning methods for comparing the performance of the deep learning methods with the state-of-the-art machine learning methods. And then we compare the performance of the reviewed methods in different datasets.

1. Discriminative Models

Discriminative deep learning models classify input data into predefined labels by adaptively learning features through nonlinear transformations and probabilistic prediction. These techniques enable both extraction and classification. Typical discriminative models include CNN, recurrent neural networks (RNN) including gated recurrent unit (GRU) and long short-term memory network (LSTM), multi-layer perceptron neural network (MLPNN), and extreme learning machine (ELM).

In the reviewed studies, CNN takes a proportion of 42%, which is in the dominant place. And other methods like LSTM(9%),RNN(2%), MLPNN(7%) also take some percentage of the studies. It shows that the discriminative models are still taking the dominant place with the percentage about 60%.

1. Representative Models

Representative deep learning models extract features unsupervised and are versatile for tasks such as clustering and classification. Typical representative DL models include deep autoencoders (DAEs), deep restricted Boltzmann machines (D-RBMs), and deep belief networks (DBN).

In the reviewed studies, representative models also take some percentage, like DAE(3%), DBN(2%) but still very small.

1. Generative Models

Generative models are deep learning architectures that learn the joint probability distribution of input data and target labels. As a result, they facilitate the quality and quantity of the training data and are primarily utilized in generating training samples and data augmentation. The two most popular types of generative deep learning models are GANs and VAEs.

But there is little research in the reviewed studies use this category of method.

#Different methods in hybrid models are counted separately

##Others means the method does not include any of the method in the figure

In ASU DB[4] dataset, the method of DTCWT+CNN with GRU[82] combining the feature extraction method of DTCWT and deep learning method CNN and GRU reach an average overall accuracy of 98.83%, which is nearly 100%. And some other methods like CSP+CNN+RNN+DAE+MLPNN(90.7%)[17], CNN/LSTM(93.5%)[18], CNN(92.5%)[21], SPWVD+CNN(92.07%)[87], Adaptive RDWT+Bagging(91.04%)[89], Bandwise-CSP(db43-wavelet)+SVM-RBF(94%)[92] are all above 90% accuracy.

In BCI Competition V-3[6] dataset, Spectro-Spatio-Temporal EEG Representation+CNN[96] reaches the highest accuracy of 70.19±7.29%, and except Riemannian geometry and Covariate Estimation+KNN[94] reaches an accuracy of 70% there isn’t any other methods achieving accuracy over 70%.

In Coretto DB [3], DWT+CNN [39] and TCNN with CNN[109] both reach an accuracy of 96.49%, but there isn’t any other methods achieving accuracy over 90%. And many of the methods are with low accuracies between 20%-40%.

In FEIS [7], DWT+ResNet152V2[122] reaches a highest accuracy of 89.01%, LPC+Integrated Stacking Classifier(SVM,CNN,MLP)[119] reaches 86.2% and Time-frequency and Statistical Characteristics+CNN[121] reaches 80%.

In Kara One [2], STFT+Multi-channel Convolution Neural Network[79] reaches the highest accuracy of 93.8%. And there are also CWT+CNN[49], DTCWT+CNN with GRU[64], DTCWT with GAF+DenseNet[73], Correntropy Spectral Density+KNN[75] and Statistical Characteristics with ANOVA+CapsNet[77] reach accuracies of 90.68%, 91.51%, 90.68%, 90.25% and 90%.

In Nieto DB [5], the method with the highest accuracy is Morlet Wavelets with Inter-trial Coherence+SVM[116] with only 67.5% accuracy. All other methods are about 30% acuuracies.

TABLE II

Imagined Speech Studies Using Public Datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Architecture | Accuracy | | Reference  (%) |
| ASU DB [4] | CSP+CNN+RNN+DAE+MLPNN | | 90.7 | [17] |
| CNN/LSTM | | 93.5 | [18] |
| CSP+DWT+MLPNN | | 71.8±8.6 | [19] |
| CSP+DWT+MLPNN | | 86.2±8.7 | [20] |
| CNN | | 92.5 | [21] |
| Statistic+BiRNN | | 61.02 | [22] |
| CWT+STFT+CNN | | 89.11 | [23] |
| CNN | | 72 ± 0.08 | [24] |
| PSD+CNN | | 41.43±0.9 | [25] |
| DTCWT+CNN with GRU | | 98.83 | [82] |
| DWT(Daubechies-4)+DNN | | 73.5±7.8 | [83] |
| CSP+LSTM | | 85.2 | [84] |
| Covariance Matrix Projection to Tangent Space+ANN | | 66.72 | [85] |
| MPC and Magnitude-Squared Coherence+Resnet50 | | 88.02 | [86] |
| SPWVD+CNN | | 92.07 | [87] |
| Multivariate DMD+RF | | 85.43 | [88] |
| Adaptive RDWT+Bagging | | 91.04 | [89] |
| Sequence of Codewords+Domain Adaptation | | 62.99±4.78 | [90] |
| CNN | | 80 | [91] |
| Bandwise-CSP(db43-wavelet)+SVM-RBF | | 94 | [92] |
| BCI Competition V-3 [6] | Entropy+CNN | | 48.1±3.68 | [26] |
| CNN | | 67±0.09 | [29] |
| SiameseNet+KNN | | 48.10±3.68 | [93] |
| Riemannian geometry and Covariate Estimation+KNN | | 70 | [94] |
| DWT(Daubechies-4)+SVM | | 40.64±2.45 | [95] |
| Spectro-Spatio-Temporal EEG Representation+CNN | | 70.19±7.29 | [96] |
| Length Classifier with Words CNNs | | 59.47±6.86 | [97] |
| Catagorical PCA | | 58.51 | [98] |
| Coretto DB [3] | CNN | | 35.68 | [30] |
| CNN | | 62.37 | [31] |
| CNN | | 31.4±2.73 | [32] |
| CNN | | 23.98 | [33] |
| FBCSP+CNN | | 30 | [34] |
| Entropy+CNN | | 45±3.13 | [35] |
| FFT+CNN | | 65.62 | [36] |
| CNN | | 35.2 | [37] |
| PSD+BiLSTM | | 36.1 | [38] |
| DWT+CNN | | 96.49 | [39] |
| CNN | | 81.69 | [40] |
| VMD+GoogleNet | | 30 | [99] |
| Multivariate Fast and Adaptive EMD+ANN | | 21.53±1.33 | [100] |
| SiameseNet+KNN | | 31.40±2.73 | [101] |
| CNN with Transfer Learning | | 24.8 | [102] |
| SiameseNet+KNN | | 45.00±3.13 | [103] |
| Deep Reinforcement Learning(DQN) | | 81.69 | [104] |
| EEGNet | | 30.25 | [105] |
| BiLSTM | | 25.1 | [106] |
| TCNN with CNN | | 96.49 | [107] |
| Multivariate Signal Decompositio+XGBoost | | 51.47 | [108] |
| CNN with Transfer Learning | | 35.68 | [109] |
| APIT-MEMD+CNN with Voting | | 62.76 | [110] |
| CNN with PCA+LDA | | 29.04 | [111] |
| Statistical Time Features+KNN | | 25.58 | [112] |
| DaSalla DB [1] | CSP+CNN | | 89.3 | [41] |
| CSP+CNN | | 93.32±4 | [123] |
| FEIS [7] | CNN | | 69±13.2 | [42] |
| FFT+LSTM | | 68.64 | [43] |
| LPC+Integrated Stacking Classifier(SVM,CNN,MLP) | | 86.2 | [119] |
| Discrete Fourier Transform+SVM | | 62.4 | [120] |
| Time-frequency and Statistical Characteristics+CNN | | 80 | [121] |
| DWT+ResNet152V2 | | 89.01 | [122] |
| Kara One [2] | Entropy+DBN | | 55.4 | [44] |
| DBN-RBM | | 69.8 | [45] |
| CNN+TCN+DAE | | 53.36 | [46] |
| CSP+MLPNN | | 57.15 | [47] |
| CSP+CNN+LSTM+DAE | | 89.3 | [48] |
| CWT+CNN | | 90.68 | [49] |
| CNN | | 49.0±6.1 | [50] |
| MLPNN | | 89.39 | [51] |
| CNN | | 39±3 | [52] |
| FFT+CNN+LSTM | | 43.98 ± 0.4 | [54] |
| FFT+CNN | | 37 | [55] |
| STFT+CNN | | 87.49 | [56] |
| FFT+LSTM | | 70.2 | [57] |
| MFSC+Integrated Stacking Classifier(SVM, CNN, MLP) | | 74.92 | [63] |
| DTCWT+CNN with GRU | | 91.51 | [64] |
| Differential Characteristics+Gaussian Process Regression | | 70 | [65] |
| MPC+ANN | | 75 | [66] |
| MFCC/LPC+CNN | | 38.33±3 | [67] |
| CorrNet+Gaussian Mixture Model | | 35.81 | [68] |
| DWT+ANN | | 57.15 | [69] |
| Time-frequency and Statistical Characteristics+SVM | | 75 | [70] |
| DWT+Pretrained DenseNet201 | | 82.35 | [71] |
| Cross-Covariate+CNN-LSTM-DAE | | 77.9 | [72] |
| DTCWT with GAF+DenseNet | | 90.68 | [73] |
| MFCC/Sequency Mapped Real Transform+ANN | | 77.37 | [74] |
| Correntropy Spectral Density+KNN | | 90.25 | [75] |
| EM-CSP+Ensemble Stacking Learning | | 86.42 | [76] |
| Statistical Characteristics with ANOVA+CapsNet | | 90 | [77] |
| Spectral Density of Correntropy+SVM | | 85.30±2.4 | [78] |
| STFT+Multi-channel Convolution Neural Network | | 93.8 | [79] |
| DWT/WPD/DWPD+MLP | | 77.97 | [80] |
| Sarmiento DB [11] | FFT+CNN | | 85.66 | [58] |
| Nieto DB [5] | Time-frequency and Statistical Characteristics+SVM | | 33.9 | [113] |
| Siamese VAE with Domain Generalisation | | 29.92 | [114] |
| CNN | | 29.67 | [115] |
| Morlet Wavelets with Inter-trial Coherence+SVM | | 67.5 | [116] |
| BiLSTM | | 36.1 | [117] |
| Shift-Invariant Sparse Coding+CNN | | 30.7 | [118] |
| Subject-Independent Meta-Learning: 50 Target Trials | | 31.15±2.16 | [124] |

IV. Discussion

The review above discloses papers of BCI with EEG and Deep Learning in the past five years. As we can see, most of the papers above use CNN, LSTM, MLPNN, Autoencoder and other popular deep learning methods. In the high accuracy methods, the most used feature extraction methods are CSP, CWT and DWT, and deep learning models often comprise of CNN and RNN (GRU, LSTM). In these high accuracy methods, directly using time-series raw data is the minority, which shows the significance of feature extraction. But there are also feature extraction using neural networks, we cannot distinct them from methods without feature extraction anyway, which needs concise analysis of these methods. Combining these methods has the potential of reaching higher accuracy imagined speech BCI. It’s worth further exploration of these methods. Some of them also focus on specific features related to EEG or neuroscience, but few of the methods are specifically designed for BCI with EEG or even for imagined speech. With more specification, we can design methods better than these normal and non-specific deep learning methods, which can better fit the demand of BCI. The future study of BCI should focus more on special features of the brain which may give inspiration to the development of BCI. And other methods except nowadays popular discriminative methods like representative methods and generative methods are neglected, there’s little research studying these kinds of methods and we need more effort to exploring the performance of these kinds of methods which have the potential of finding latent representation of EEG signals or directly generating voice from EEG data.

We can also see the average performance difference between different datasets. That’s maybe because of experimental paradigms, device, channel number, channel position, class number or simply the problem of their methods. Testing a method in different datasets is necessary in future studies to show the generalization of the problem. Also, most of the datasets used now contain too few words or phonemes. There hasn’t been any dataset containing all phonemes or even words (which is too huge to collect, so recording phoneme data will be more practical). Only with enough data, then we can train generalizable enough models. The data for now is far from enough, we should also do more work on data collection which will help in future BCI study. We can also combine different datasets together to get more data and it can also help improve generalization in different environments, devices, subjects. The future collection of data should be mainly phonemes because of its limited and acceptable number rather than words or even sentences that will be expensive to collect because of their huge amount. With the combination of phonemes, we can get infinite number of pronunciations of words then we can reach practical BCI system generation voice. And we can also divide imagined words into phonemes to get more data, but the method needs to be developed. And then we can develop methods detecting phonemes and combine them in time series to realize imagined speech dataset. Many of the paradigms of these datasets are only with covert speech, there’s only some of them have overt speech EEG data. Studies show that there are some differences in brain activity between covertly speaking and overtly speaking. When imagining speech, there might be some difference in the way the subjects imagine speech. So, it’s still worth studying the performance of transforming EEG data into voice. Another solution is that the subjects can listen to a recorded sound or speech and repeat in the following imagining speech process, this can standardize the way subjects imagine speech and improve the accuracy and even can transform EEG data into any kind of sound or voice. This kind of new paradigm is also worth exploring the effect.

Nowadays, the use of BCI in many areas like healthcare, entertainment, gaming, communication, control, research and so on have been done many studies. All the mentioned areas need highly accurately recognizing many kinds of signals. Despite the demand for online response time for BCI considering most of the studies above are offline, many of the important papers above still use binary classification and the methods using multi-class classification are with low accuracy which cannot satisfy the demand of the usages above. Combining datasets can also increase the number of classes but it needs to be further studied. The road to BCI that can be used in real life circumstances is still a long way to go.

V. Conclusion

In conclusion, this article shows the recent research condition of the imagined speech BCI. We found that though many up-to-date deep learning methods are used in these studies. In the high accuracy methods, the most used feature extraction methods are CSP, CWT and DWT, and deep learning models often comprise of CNN and RNN (GRU, LSTM). But other methods like representative methods and generative methods need more study. The study of imagined speech BCI is still in primary stage for shortage of data and high-accuracy BCI multi-classification algorithms, we can solve it by combining the datasets. Using phonemes rather than words and standardizing imagination content are more practical choices for paradigms. We should develop more BCI-specific algorithms, focusing on more brain features, to make imagined speech BCI applicable in real-life and a more connected and convenient world. The road to BCI that can be used in real life circumstances is still a long way to go.

References

[1] DaSalla, Charles S., et al. "Single-trial classification of vowel speech imagery using common spatial patterns." Neural networks 22.9 (2009): 1334-1339.

[2] S. Zhao and F. Rudzicz, “Classifying phonological categories in imagined and articulated speech,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2015, pp. 992–996.

[3] G. A. Pressel Coretto, I. E. Gareis, and H. L. Rufiner, “Open access database of EEG signals recorded during imagined speech,” in SPIE Proc., Jan. 2017, pp. 1–26.

[4] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, “Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features,” J. Neural Eng., vol. 15, no. 1, Feb. 2018, Art. no. 016002.

[5] N. Nieto, V. Peterson, H. L. Rufiner, J. E. Kamienkowski, and R. Spies, “Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition,” Sci. Data, vol. 9, no. 1, p. 52, Feb. 2022.

[6] Int. BCI Competition. (2020). 2020 International BCI Competition. [Online]. Available: <https://osf.io/pq7vb/>

[7] S. Wellington and J. Clayton, “Fourteen-channel EEG with imagined speech (FEIS) dataset,” Univ. Edinburgh, Edinburgh, Scotland, 2019, doi: 10.5281/zenodo.3369178.

[8] F. Simistira Liwicki et al., “Bimodal electroencephalography-functional magnetic resonance imaging dataset for inner-speech recognition,” Scientific Data, vol. 10, no. 1, p. 378, Jun. 2023.

[9] B. Dekker, A. C. Schouten, and O. Scharenborg, “DAIS: The Delft database of EEG recordings of Dutch articulated and imagined speech,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Jun. 2023, pp. 1–5.

[10] J. LaRocco, Q. Tahmina, S. Lecian, J. Moore, C. Helbig, and S. Gupta, “Evaluation of an English language phonemebased imagined speech brain computer interface with low-cost electroencephalography,” Frontiers Neuroinform., vol. 17, Dec. 2023, Art. no. 1306277.

[11] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento, and J. B. Rodríguez, “Recognition of EEG signals from imagined vowels using deep learning methods,” Sensors, vol. 21, no. 19, p. 6503, Sep. 2021.

[12] A. Jahangiri and F. Sepulveda, “The relative contribution of highgamma linguistic processing stages of word production, and motor imagery of articulation in class separability of covert speech tasks in EEG data,” J. Med. Syst., vol. 43, no. 2, p. 20, Feb. 2019.

[13] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, “Envisioned speech recognition using EEG sensors,” Pers. Ubiquitous Comput., vol. 22, no. 1, pp. 185–199, Feb. 2018.

[14] Zhang, Liying, et al. "Speech imagery decoding using EEG signals and deep learning: A survey." IEEE Transactions on Cognitive and Developmental Systems (2024).

[15] Tang, Jiye, et al. "Imagined Speech Reconstruction from Neural Signals–An Overview of Sources and Methods." IEEE Transactions on Instrumentation and Measurement (2024).

[16] Rahman, Nimra, et al. "Advances in brain-computer interface for decoding speech imagery from EEG signals: a systematic review." Cognitive Neurodynamics (2024): 1-19.

[17] P. Saha and S. Fels, “ Hierarchical deep feature learning for decoding

imagined speech from EEG,” in Proc. AAAI Conf. Artif. Intell., vol. 33,

2019, pp. 10019-10020.

[18] M. arhi and A. . Tewfik, “Classifying imaginary vowels from

frontal lobe eeg via deep learning,” in Proc. 28th Eur. Signal Process.

Conf. (EUSIPCO), Jan. 2021, pp. 1195-1199.

[19] J. T. anachakel, A. Ramakrishnan, and T. Ananthapadmanabha, “A

novel deep learning architecture for decoding imagined speech from

EEG,” . [Online]. Available: arXiv: 2003.09374.

[20] J. T. Panachakel, R. A. Ganesan, and T. Ananthapadmanabha,

“Common Spatial attern Based Data Augmentation Technique for

Decoding magined Speech,” in Proc. IEEE Int. Conf. Electron.,

Comput. Commun. Technol. (CONECCT), Jul. 2021, pp. 1-5.

[21] J. T. anachakel and R. A. Ganesan, “Decoding imagined speech from

EEG using transfer learning,” IEEE Access, vol. 9, pp. 135371-135383,

Oct. 2021.

[22] M. Jiménez-Guarneros and P. Gómez-Gil, “Standardization-refinement

domain adaptation method for cross-subject EEG-based classification

in imagined speech recognition,” Pattern Recognit. Lett., vol. 141, pp.

54-60, Jan. 2021.

[23] A. Kamble, . . Ghare, and . Kumar, “Deep-learning-based BCI for

automatic imagined speech recognition using S W D,” IEEE Trans.

Instrum. Meas., vol. 72, pp. 1-10, Jan. 2023.

[24] H.-J. Ahn, D.-H. Lee, J.-H. Jeong, and S.-W. Lee, “Multiscale

Convolutional Transformer for EEG Classification of Mental Imagery in

Different Modalities,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31,

pp. 646–656, Feb. 2023.

[25] J. S. García-Salinas, A. A. Torres-García, C. A. Reyes-Garćia, and L.

Villaseñor- ineda, “ ntra-subject class-incremental deep learning

approach for EEG-based imagined speech recognition,” Biomed. Signal

Process. Control, vol. 81, Mar. 2023, Art. no. 104433.

[26] D.-Y. Lee, M. Lee, and S.-W. Lee, “Decoding imagined speech based

on deep metric learning for intuitive BC communication,” IEEE Trans.

Neural Syst. Rehabil. Eng., vol. 29, pp. 1363-1374, Jul. 2021.

[27] W. Ko, E. Jeon, and H.- . Suk, “Spectro-Spatio-Temporal EEG Repres-

entation Learning for magined Speech Recognition,” in Proc. 6th

IAPR Asian Conf. Pattern Recognit. (ACPR), Nov. 2021, pp. 335-346.

[28] B.-H. Lee, B.-H. Kwon, D.-Y. Lee, and J.- . Jeong, “Speech magery

Classification using Length-Wise Training based on Deep Learning,”

in Proc. 9th Int. Winter Conf. Brain-Comput. Interface (BCI), Feb.

2021, pp. 1-5.

[29] W. Ko, E. Jeon, S. Jeong, and H.- . Suk, “Multi-scale neural network

for EEG representation learning in BC ,” IEEE Comput. Intell. Mag.,

vol. 16, no. 2, pp. 31–45, May 2021.

[30] C. Cooney, A. Korik, R. Folli, and D. Coyle, “Evaluation of hyperpara-

meter optimization in machine and deep learning methods for

decodingimagined speech EEG,” Sensors, vol. 20, no. 16, p. 4629,

2020.

[31] C. Cooney, A. Korik, F. Raffaella, and D. Coyle, “Classification of

imagined spoken word-pairs using convolutional neural networks,” in

Proc. Graz BCI Conf., Sep. 2019, pp. 338–343.

[32] D.-Y. Lee, M. Lee, and S.-W. Lee, “Classification of imagined speech

using Ng neural network,” in Proc. IEEE Int. Conf. Syst., Man,

Cybern. (SMC), Oct. 2020, pp. 2979–2984.

[33] M.-O. Tamm, . Muhammad, and N. Muhammad, “Classification of

vowels from imagined speech with convolutional neural networks,”

Computers, vol. 9, no. 2, p. 46, Jun. 2020.

[34] C. Cooney, A. Korik, R. Folli, and D. Coyle, “Evaluation of hyperpara-

meter optimization in machine and deep learning methods for

decodingimagined speech EEG,” Sensors, vol. 20, no. 16, p. 4629,

2020.

[35] D.-Y. Lee, M. Lee, and S.-W. Lee, “Decoding imagined speech based

on deep metric learning for intuitive BC communication,” IEEE Trans.

Neural Syst. Rehabil. Eng., vol. 29, pp. 1363-1374, Jul. 2021.

[36] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento,

and J. B. Rodríguez, “Recognition of EEG signals from imagined

vowels using deep learning methods,” Sensors, vol. 21, no. 19, p. 6503,

2021.

[37] F. Simistira Liwicki, V. Gupta, R. Saini, K. De, and M. Liwicki,

“Rethinking the Methods and Algorithms for nner Speech Decoding

and Making Them Reproducible,” NeuroSci, vol. 3, no. 2, pp. 226-244,

Apr. 2022.

[38] F. Gasparini, E. Cazzaniga, and A. Saibene, “ nner speech recognition

through electroencephalographic signals,” . [Online]. Available:

arXiv: 2210.06472.

[39] N. C. Mahapatra and . Bhuyan, “Multiclass Classification of magined

Speech Vowels and Words of Electroencephalography Signals Using

Deep Learning,” Adv. Hum. -Comput. Interact., vol. 2022, 2022.

[40] N. C. Mahapatra and . Bhuyan, “Decoding of magined Speech

Neural EEG Signals sing Deep Reinforcement Learning Technique,”

in Proc. IEEE Int. Conf. Advancements Smart, Secure intell. Comput.

(ASSIC), Nov. 2022, pp. 1-6.

[41] J. M. Macías-Macías, J. A. Ramírez-Quintana, G. Ramírez-Alonso, and

M. I. Chacón-Murguía, “Deep learning networks for vowel speech

imagery,” in Proc. 17th Int. Conf. Electr. Eng., Comput. Sci. Autom.

Control (CCE), Nov. 2020, pp. 1-6.

[42] J. Clayton, S. Wellington, C. Valentini-Botinhao, and O. Watts,

“Decoding magined, eard, and Spoken Speech: Classification and

Regression of EEG Using a 14-Channel Dry-Contact Mobile eadset,”

in Proc. Interspeech, Oct. 2020, pp. 4886-4890.

[43] P. Agarwal and S. Kumar, “Electroencephalography‐based imagined

speech recognition using deep long short ‐ term memory network,”

ETRI J., vol. 44, no. 4, pp. 672-685, Jun. 2022.

[44] S. Zhao and F. Rudzicz, “Classifying phonological categories in

imagined and articulated speech,” in Proc. IEEE Int. Conf. Acoust.,

Speech Signal Process. (ICASSP), Apr. 2015, pp. 992–996.

[45] P. Sun and J. Qin, “Neural networks based eeg-speech models,” 6.

[Online]. Available: arXiv: 1612.05369.

[46] P. Saha, M. Abdul-Mageed, and S. Fels, “Speak your mind! Towards

imagined speech recognition with hierarchical deep learning,” in Proc.

Interspeech, Graz, Austria, Sep. 2019, pp. 141–145.

[47] J. T. Panachakel, A. Ramakrishnan, and T. Ananthapadmanabha,

“Decoding imagined speech using wavelet features and deep neural

networks,” in Proc. IEEE 16th India Council Int. Conf. (INDICON),

Rajkot, India, Dec. 2019, pp. 1-4.

[48] P. Saha, S. Fels, and M. Abdul-Mageed, “Deep learning the EEG

manifold for phonological categorization from active thoughts,” in

Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May

2019, pp. 2762–2766.

[49] M. M. slam and M. M. . Shuvo, “DenseNet based speech imagery

EEG signal classification using Gramian Angular Field,” in Proc. 5th

Int. Conf. Adv. Electr. Eng. (ICAEE), Sep. 2019, pp. 149-154.

[50] J. Clayton, S. Wellington, C. Valentini-Botinhao, and O. Watts,

“Decoding magined, eard, and Spoken Speech: Classification and

Regression of EEG Using a 14-Channel Dry-Contact Mobile eadset,”

in Proc. Interspeech, Oct. 2020, pp. 4886-4890.

[51] R. A. Sharon and . A. Murthy, “Correlation based Multi-phasal

models for improved imagined speech EEG recognition,” .

[Online]. Available: arXiv: 2011.02195.

[52] A.-L. Rusnac and O. Grigore, “Convolutional Neural Network applied

in EEG imagined phoneme recognition system,” in Proc. 12th Int.

Symp. Adv. Topics Electr. Eng. (ATEE), Mar. 2021, pp. 1-4.

[53] S. Datta and N. . Boulgouris, “Recognition of grammatical class of

imagined words from EEG signals using convolutional neural network,”

Neurocomputing, vol. 465, pp. 301-309, Nov. 2021.

[54] A.-L. Rusnac and O. Grigore, “ Imaginary Speech Recognition sing a

Convolutional Network with Long-Short Memory,” Appl. Sci., vol. 12,

no. 22, Nov. 2022, Art. no. 11873.

[55] A.-L. Rusnac and O. Grigore, “CNN Architectures and Feature

Extraction Methods for EEG Imaginary Speech Recognition,” Sensors,

vol. 22, no. 13, p. 4679, 2022.

[56] Q. eting and G. Nuo, “Research on the Classification Algorithm of

maginary Speech EEG Signals Based on Twin Neural Network,” in

Proc. IEEE 7th Int. Conf. Signal Image Process. (ICSIP), Jul. 2022, pp.

211-216.

[57] N. Kobayashi and T. Morooka, “Application of igh-accuracy Silent

Speech BC to Biometrics using Deep Learning,” in Proc. 9th Int.

Winter Conf. Brain-Comput. Interface (BCI), Feb. 2021, pp. 1-6.

[58] F. Simistira Liwicki, V. Gupta, R. Saini, K. De, and M. Liwicki,

“Rethinking the Methods and Algorithms for nner Speech Decoding

and Making Them Reproducible,” NeuroSci, vol. 3, no. 2, pp. 226-244,

Apr. 2022.

[59] B. van den Berg, S. van Donkelaar, and M. Alimardani, “ nner speech

classification using eeg signals: A deep learning approach,” in Proc.

IEEE 2nd Int. Conf. Hum.-Mach. Syst. (ICHMS), Magdeburg, Germany,

Sep. 2021, pp. 1-4.

[60] B. van den Berg, S. van Donkelaar, and M. Alimardani, “ nner speech

classification using eeg signals: A deep learning approach,” in Proc.

IEEE 2nd Int. Conf. Hum.-Mach. Syst. (ICHMS), Magdeburg, Germany,

Sep. 2021, pp. 1-4.

[61] F. Gasparini, E. Cazzaniga, and A. Saibene, “ nner speech recognition

through electroencephalographic signals,” . [Online]. Available:

arXiv: 2210.06472.

[62] F. Gasparini, E. Cazzaniga, and A. Saibene, “ nner speech recognition

through electroencephalographic signals,” . [Online]. Available:

arXiv: 2210.06472.

[63] N. Ramkumar and D. K. Renuka, “A novel BCI-based silent speech

recognition using hybrid feature extraction techniques and integrated

stacking classifier: A novel BCI-based silent speech recognition,” J. Sci.

Ind. Res. (JSIR), vol. 82, no. 11, pp. 1165–1176, 2023.

[64] A. Hernandez-Galvan, G. Ramirez-Alonso, and J. Ramirez-Quintana,

“A prototypical network for few-shot recognition of speech imagery

data,” Biomed. Signal Process. Control, vol. 86, Sep. 2023,

Art. no. 105154.

[65] H. Wu and F. Chen, “A temporal envelope-based speech reconstruction

approach with EEG signals during speech imagery,” in Proc. Asia–

Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC),

Dec. 2020, pp. 894–899.

[66] J. T. Panachakel and R. A. G, “Classification of phonological categories

in imagined speech using phase synchronization measure,” in Proc.

43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Nov. 2021,

pp. 2226–2229.

[67] A.-L. Rusnac and O. Grigore, “Convolutional neural network applied in

EEG imagined phoneme recognition system,” in Proc. 12th Int. Symp.

Adv. Topics Electr. Eng. (ATEE), Mar. 2021, pp. 1–4.

[68] R. A Sharon and H. A Murthy, “Correlation based multi-phasal

models for improved imagined speech EEG recognition,” 2020,

arXiv:2011.02195.

[69] J. T. Panachakel, A. G. Ramakrishnan, and T. V. Ananthapadmanabha,

“Decoding imagined speech using wavelet features and deep neural

networks,” in Proc. IEEE 16th India Council Int. Conf. (INDICON),

Dec. 2019, pp. 1–4.

[70] J. Clayton, S. Wellington, C. Valentini-Botinhao, and O. Watts,

“Decoding imagined, heard, and spoken speech: Classification and

regression of EEG using a 14-channel dry-contact mobile headset,”

in Proc. Interspeech, Oct. 2020, pp. 4886–4890.

[71] N. C. Mahapatra and P. Bhuyan, “Decoding of imagined speech

electroencephalography neural signals using transfer learning method,”

J. Phys. Commun., vol. 7, no. 9, Sep. 2023, Art. no. 095002.

[72] P. Saha, S. Fels, and M. Abdul-Mageed, “Deep learning the EEG

manifold for phonological categorization from active thoughts,” in

Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP),

May 2019, pp. 2762–2766.

[73] M. M. Islam and M. M. H. Shuvo, “DenseNet based speech imagery

EEG signal classification using gramian angular field,” in Proc. 5th

Int. Conf. Adv. Electr. Eng. (ICAEE), Sep. 2019, pp. 149–154, doi:

10.1109/ICAEE48663.2019.8975572.

[74] P. P. Mini, T. Thomas, and R. Gopikakumari, “EEG based direct speech

BCI system using a fusion of SMRT and MFCC/LPCC features with

ANN classifier,” Biomed. Signal Process. Control, vol. 68, Jul. 2021,

Art. no. 102625.

[75] M. A. Bakhshali, M. Khademi, A. Ebrahimi-Moghadam, and S.

Moghimi, “EEG signal classification of imagined speech based on

Riemannian distance of correntropy spectral density,” Biomed. Signal

Process. Control, vol. 59, May 2020, Art. no. 101899.

[76] D. Alizadeh and H. Omranpour, “EM-CSP: An efficient multiclass

common spatial pattern feature method for speech imagery EEG signals

recognition,” Biomed. Signal Process. Control, vol. 84, Jul. 2023,

Art. no. 104933.

[77] J. M. Macías-Macías, J. A. Ramírez-Quintana, M. I. Chacón-Murguía,

A. A. Torres-García, and L. F. Corral-Martínez, “Interpretation of a

deep analysis of speech imagery features extracted by a capsule neural

network,” Comput. Biol. Med., vol. 159, Jun. 2023, Art. no. 106909.

[78] M. A. Bakhshali, M. Khademi, and A. Ebrahimi-Moghadam,

“Investigating the neural correlates of imagined speech: An EEG-based

connectivity analysis,” Digit. Signal Process., vol. 123, Apr. 2022,

Art. no. 103435.

[79] S. Datta and N. V. Boulgouris, “Recognition of grammatical class of

imagined words from EEG signals using convolutional neural network,”

Neurocomputing, vol. 465, pp. 301–309, Nov. 2021.

[80] M. P. P., T. Thomas, and R. Gopikakumari, “Wavelet feature selection

of audio and imagined/vocalized EEG signals for ANN based

multimodal ASR system,” Biomed. Signal Process. Control, vol. 63,

Jan. 2021, Art. no. 102218.

[81] P. Saha, M. Abdul-Mageed, and S. Fels, “Speak your mind! Towards

imagined speech recognition with hierarchical deep learning,” 2019,

arXiv:1904.05746.

[82] A. Hernandez-Galvan, G. Ramirez-Alonso, and J. Ramirez-Quintana,

“A prototypical network for few-shot recognition of speech imagery

data,” Biomed. Signal Process. Control, vol. 86, Sep. 2023,

Art. no. 105154.

[83] J. T. Panachakel, R. A. Ganesan, and T. Ananthapadmanabha,

“Common spatial pattern based data augmentation technique for

decoding imagined speech,” in Proc. IEEE Int. Conf. Electron.,

Comput. Commun. Technol. (CONECCT), Jul. 2021, pp. 1–5.

[84] J. T. Panachakel and A. Ramakrishnan, “DCLL—A deep network for

possible real-time decoding of imagined words,” in Proc. Int. Symp.

Intell. Informat., 2022, pp. 3–12.

[85] A. Singh and A. Gumaste, “Decoding imagined speech and computer

control using brain waves,” J. Neurosci. Methods, vol. 358, Jul. 2021,

Art. no. 109196.

[86] J. T. Panachakel and R. A. Ganesan, “Decoding imagined

speech from EEG using transfer learning,” IEEE Access, vol. 9,

pp. 135371–135383, 2021.

[87] A. Kamble, P. H. Ghare, and V. Kumar, “Deep-learning-based BCI for

automatic imagined speech recognition using SPWVD,” IEEE Trans.

Instrum. Meas., vol. 72, pp. 1–10, 2023.

[88] A. Siva Sankar Reddy and R. Bilas Pachori, “Multivariate dynamic

mode decomposition for automatic imagined speech recognition using

multichannel EEG signals,” IEEE Sensors Lett., vol. 8, no. 2, pp. 1–4,

Feb. 2024.

[89] A. Kamble, P. H. Ghare, and V. Kumar, “Optimized rational dilation

wavelet transform for automatic imagined speech recognition,” IEEE

Trans. Instrum. Meas., vol. 72, pp. 1–10, 2023.

[90] M. Jiménez-Guarneros and P. Gómez-Gil, “Standardization-refinement

domain adaptation method for cross-subject EEG-based classification

in imagined speech recognition,” Pattern Recognit. Lett., vol. 141,

pp. 54–60, Jan. 2021, doi: 10.1016/j.patrec.2020.11.013.

[91] S. Liu and J. Chan, “Testing the effectiveness of CNN and GNN and

exploring the influence of different channels on decoding covert speech

from EEG signals,” in Proc. 12th Int. Conf. Comput. Syst.-Biol. Bioinf.,

vol. 10, Oct. 2021, pp. 17–26.

[92] S. Biswas and R. Sinha, “Wavelet filterbank-based EEG rhythmspecific spatial features for covert speech classification,” IET Signal

Process., vol. 16, no. 1, pp. 92–105, Feb. 2022.

[93] D.-Y. Lee, M. Lee, and S.-W. Lee, “Decoding imagined speech based

on deep metric learning for intuitive BCI communication,” IEEE Trans.

Neural Syst. Rehabil. Eng., vol. 29, pp. 1363–1374, 2021.

[94] F. P. Kalaganis, K. Georgiadis, V. P. Oikonomou, S. Nikolopoulos,

N. A. Laskaris, and I. Kompatsiaris, “Exploiting approximate

joint diagonalization for covariance estimation in imagined speech

decoding,” in Proc. Int. Conf. Brain Informat., 2023, pp. 409–419.

[95] D. Pawar and S. Dhage, “Imagined speech classification using EEG

based brain-computer interface,” in Proc. IEEE 11th Int. Conf.

Commun. Syst. Netw. Technol. (CSNT), Apr. 2022, pp. 662–666.

[96] W. Ko, E. Jeon, and H.-I. Suk, “Spectro-spatio-temporal EEG

representation learning for imagined speech recognition,” in Proc.

Asian Conf. Pattern Recognit., 2021, pp. 335–346.

[97] B.-H. Lee, B.-H. Kwon, D.-Y. Lee, and J.-H. Jeong, “Speech imagery

classification using length-wise training based on deep learning,” in

Proc. 9th Int. Winter Conf. Brain-Comput. Interface (BCI), Feb. 2021,

pp. 1–5.

[98] X.-B. Zheng, B. Wing-Kuen Ling, S.-Y. Zheng, and C.-J. Li,

“Supervised categorized principal component analysis for imagined

speech classification via applying singular value decomposition on a

symmetry matrix,” Biomed. Signal Process. Control, vol. 86, Sep. 2023,

Art. no. 105324.

[99] P. Ray, S. S. Reddy, and T. Banerjee, “Various dimension reduction

techniques for high dimensional data analysis: A review,” Artif. Intell.

Rev., vol. 54, no. 5, pp. 3473–3515, Jun. 2021.

[100] S. Dash, R. K. Tripathy, G. Panda, and R. B. Pachori, “Automated

recognition of imagined commands from EEG signals using multivariate fast and adaptive empirical mode decomposition based method,”

IEEE Sensors Lett., vol. 6, no. 2, pp. 1–4, Feb. 2022.

[101] D.-Y. Lee, M. Lee, and S.-W. Lee, “Classification of imagined speech

using Siamese neural network,” in Proc. IEEE Int. Conf. Syst. Man,

Cybern. (SMC), Oct. 2020, pp. 2979–2984.

[102] M.-O. Tamm, Y. Muhammad, and N. Muhammad, “Classification of

vowels from imagined speech with convolutional neural networks,”

Computers, vol. 9, no. 2, p. 46, Jun. 2020.

[103] D.-Y. Lee, M. Lee, and S.-W. Lee, “Decoding imagined speech based

on deep metric learning for intuitive BCI communication,” IEEE Trans.

Neural Syst. Rehabil. Eng., vol. 29, pp. 1363–1374, 2021.

[104] N. C. Mahapatra and P. Bhuyan, “Decoding of imagined speech neural

EEG signals using deep reinforcement learning technique,” in Proc. Int.

Conf. Advancements Smart, Secure Intell. Comput. (ASSIC), Nov. 2022,

pp. 1–6.

[105] C. Cooney, A. Korik, R. Folli, and D. Coyle, “Evaluation of

hyperparameter optimization in machine and deep learning methods

for decoding imagined speech EEG,” Sensors, vol. 20, no. 16, p. 4629,

Aug. 2020.

[106] F. Gasparini, E. Cazzaniga, and A. Saibene, “Inner speech recognition

through electroencephalographic signals,” 2022, arXiv:2210.06472.

[107] N. C. Mahapatra and P. Bhuyan, “Multiclass classification of imagined

speech vowels and words of electroencephalography signals using deep

learning,” Adv. Hum.-Comput. Interact., vol. 2022, pp. 1–10, Jul. 2022.

[108] S. Dash, R. K. Tripathy, D. K. Dash, G. Panda, and R. B. Pachori,

“Multiscale domain gradient boosting models for the automated

recognition of imagined vowels using multichannel EEG signals,” IEEE

Sensors Lett., vol. 6, no. 11, pp. 1–4, Nov. 2022.

[109] C. Cooney, R. Folli, and D. Coyle, “Optimizing layers improves CNN

generalization and transfer learning for imagined speech decoding from

EEG,” in Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC), Oct. 2019,

pp. 1311–1316.

[110] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento,

and J. B. Rodríguez, “Recognition of EEG signals from imagined

vowels using deep learning methods,” Sensors, vol. 21, no. 19, p. 6503,

Sep. 2021.

[111] O. Banerjee, D. Govind, A. K. Dubey, and S. V. Gangashetty,

“Significance of dimensionality reduction in CNN-based vowel

classification from imagined speech using electroencephalogram

signals,” in Proc. Int. Conf. Speech Comput., 2022, pp. 44–55.

[112] S. N. Liyanagoonawardena, A. T. Palihakkara, D. N. Mudalige,

C. J. U. Gamage, A. C. De Silva, and T. Chang, “Significance of

subject-dependency for imagined speech classification using EEG,”

in Proc. IEEE 17th Int. Conf. Ind. Inf. Syst. (ICIIS), Aug. 2023,

pp. 541–546.

[113] N. R. Merola, N. G. Venkataswamy, and M. H. Imtiaz, “Can

machine learning algorithms classify inner speech from EEG brain

signals?” in Proc. IEEE World AI IoT Congr. (AIIoT), Jun. 2023,

pp. 0466–0470.

[114] H. W. Ng and C. Guan, “Efficient representation learning for inner

speech domain generalization,” in Proc. Int. Conf. Comput. Anal.

Images Patterns, 2023, pp. 131–141.

[115] B. v. d. Berg, S. v. Donkelaar, and M. Alimardani, “Inner speech

classification using EEG signals: A deep learning approach,” in Proc.

IEEE 2nd Int. Conf. Human-Machine Syst. (ICHMS), Sep. 2021,

pp. 1–4.

[116] D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, “Inner

speech classification using inter-trial coherence framework for feature

extraction,” in Proc. 19th Int. Symp. Med. Inf. Process. Anal. (SIPAIM),

Nov. 2023, pp. 1–6.

[117] F. Gasparini, E. Cazzaniga, and A. Saibene, “Inner speech recognition

through electroencephalographic signals,” 2022, arXiv:2210.06472.

[118] K. Tyrrell and M. H. Kapourchali, “Unsupervised learning for exploring

hidden structures in self-talk,” in Proc. IEEE 19th Int. Conf. Body

Sensor Netw. (BSN), Oct. 2023, pp. 1–4.

[119] N. Ramkumar and D. K. Renuka, “A novel BCI-based silent speech

recognition using hybrid feature extraction techniques and integrated

stacking classifier: A novel BCI-based silent speech recognition,” J. Sci.

Ind. Res. (JSIR), vol. 82, no. 11, pp. 1165–1176, 2023.

[120] R. Sakai, A. Kai, and S. Nakagawa, “Classification of imagined and

heard speech using amplitude spectrum and relative phase of EEG,” in

Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech), Mar. 2021,

pp. 373–375.

[121] J. Clayton, S. Wellington, C. Valentini-Botinhao, and O. Watts,

“Decoding imagined, heard, and spoken speech: Classification and

regression of EEG using a 14-channel dry-contact mobile headset,”

in Proc. Interspeech, Oct. 2020, pp. 4886–4890.

[122] N. C. Mahapatra and P. Bhuyan, “Decoding of imagined speech

electroencephalography neural signals using transfer learning method,”

J. Phys. Commun., vol. 7, no. 9, Sep. 2023, Art. no. 095002.

[123] J. A. Ramirez-Quintana, J. M. Macias-Macias, G. Ramirez-Alonso, M.

I. Chacon-Murguia, and L. F. Corral-Martinez, “A novel deep capsule

neural network for vowel imagery patterns from eeg signals,” Biomed.

Signal Process. Control, vol. 81, Mar. 2023, Art. no. 104500.

[124] Ng, Han Wei, and Cuntai Guan. "Subject-independent meta-learning framework towards optimal training of eeg-based classifiers." Neural Networks 172 (2024): 106108.