Inner speech classification based on electroencephalography (EEG) signals and support vector machine (SVM)

Xiao Wei (xiao.wei@uni-due.de)
University of Duisburg-Essen
Alvin Immanuel Surjana (alvin.surjana@hs-niederrhein.de)
University of Duisburg-Essen
Dirk Söffker (soeffker@uni-due.de)
University of Duisburg-Essen

Research Article

Keywords:

DOI: https://doi.org/

RESEARCH

Inner speech classification based on electroencephalography (EEG) signals and support vector machine (SVM)

Xiao Wei*, Alvin Immanuel Surjana, and Dirk Söffker

*Correspondence:
xiao.wei@uni-due.de
Chair of Dynamics and Control,
University of Duisburg-Essen,
Duisburg, Germany
Full list of author information is available at the end of the article

Abstract

Inner speech is a form of self-directed dialogue which plays an important role in cognitive development, speech monitoring, executive function, and psychopathology. Despite of a growing knowledge on its phenomenology, development, and function, approaches to the scientific study of inner speech have remained diffuse and largely unintegrated. Electroencephalography (EEG) which is a non-intrusive approach for brain-computer interface (BCI) brings new options to inner speech studies. Due to the advantages of EEG, more and more studies are related to inner speech apply EEG signals. In this contribution, different words expressed in inner speech are distinguished by applying EEG signals and support vector machine (SVM). Electroencephalography data from 'Thinking out loud' dataset open to public are employed. In the experiment, numerous EEG data are acquired from the 128 sensors located in the headcap. Therefore, data are filtered in the first step. Afterwards, selected data are decomposed by empirical mode decomposition (EMD) into various intrinsic mode functions (IMFs). Furthermore, IMFs are transformed using Hilbert transform to check the brain wave bands suitable for distinguishing inner speech. Lastly, single or combined of IMFs are classified by support vector machine (SVM) with various kernels. When most suitable IMFs and kernels are employed, the average results for each subject scheme are: F-score: 99.24 %, accuracy: 99.24 %, and standard deviation (SD): 0.95. Best results for all subject schemes are: F-score: 99.67 %, accuracy: 99.66 %, and standard deviation (SD): 0.27. The obtained results demonstrate that the proposed approach can differentiate EEG signals from inner speech well.

Keywords: Inner speech; Electroencephalography; Support vector machine; Savitzky-Golay filter; Empirical mode decomposition

Introduction

Inner speech (IS) (silent-, covert-, speech, verbal thought) are some of the terms used to refer to the silent production of words in one's mind, or the activity of talking to oneself silence [1]. It is defined as the silent expression of conscious thought to oneself in a coherent linguistic form [2]. Though much difficult in studying inner speech, the role it plays in different cognitive abilities, including memory and executive functioning is well established. It is linked to the development of language abilities and the advanced mental abilities to which language is linked [3]. Another skill that appears linked to inner speech is silent reading. In addition, inner speech helps certain brain disorders resulting from brain stem infarcts, traumatic brain injury, cerebral palsy, stroke, and amyotrophic lateral sclerosis, limit verbal communication despite the patient being fully aware [4]. People that cannot communicate due to neurological disorder would benefit from systems that can infer internal speech directly from brain signals. People with speech deficits would benefit from a communication system that can directly infer inner speech from brain signals – allowing them to interact more naturally with the world [5]. Therefore, inner speech recognition has been proposed as an alternative communication paradigm for brain-computer interface (BCI) [6].

Brain-computer interface is a collaboration between brain and device that enables signals from the brain to direct some external activity. The interface enables a direct communication pathway between the brain and the object to be controlled [7]. In addition to its application in medical and health field, BCI technology has potential applications in military, education, and recreation fields. According to [8], BCI system consists of four basic components: signal acquisition, signals preprocessing, feature extraction, and classification. Discriminative characteristics of the improved signals are extracted in features extraction stage. Lastly, classifiers distinguish features and allow finally the guidance of device commands.

Signal acquisition is measuring brain oscillating electrical voltages which is also named brain waves [9]. Brain waves reflects the voluntary neural actions generated by user's current activity. Methods capturing brain waves can be divided into invasive and non-invasive. Invasive recording methods implant electrodes under the scalp and measure the neural activity of the brain either intracortically from within the motor cortex or on the cortical surface (electrocorticography (ECoG)) [8]. The most relevant advantage of invasive recording is providing high temporal and spatial resolution, increasing the quality of the obtained signal, and a good signal to noise ration [8]. Main downside is the utilization of invasive surgery, and the potential for scar tissue to form around the site, which may lead to potential side effects such as seizures [10]. The invasive systems are mostly used in BCI systems experiment that use monkeys according to [11]. Instead of the surgical procedure and permanent device attachment, non-invasive recording methods record brain activity from electrodes placed on the skin and scalp. In general, non-invasive are considered the safest and low-cost type of devices. Functional magnetic resonance imaging (fMRI), functional near infrared spectroscopy (fNIRS), magnetoencephalography (MEG), electroencephalogram (EEG), and stereotactic electroencephalography (sEEG) belong to the non-invasive methods [12]. Among these non-invasive recording methods, EEG is one of the most common used methods in BCI system. It is a physiological method of choice to record the electrical activity generated by the brain via electrodes placed on the scalp surface [13]. These electrodes can only capture 'weaker' human brain signals due to the obstruction of the skull.

A large number of research focuses on EEG-based BCI system [14, 15, 16, 17, 18]. Common BCI paradigms, the signal processing, and feature extraction methods are introduced in [14]. In [15, 17], commonly employed algorithms used to design BCI systems based on EEG are presented and relevant critical properties are described. The state-of-art as well as trends in EEG-based emotion recognition system research are summarized by AI-Nafjan et al. [16] focusing on emotion detection, recognition, and classification. Brain-computer interface paradigms, EEG feature types, classification algorithms, and target applications from 2007 to 2011 are revealed in [18].

In other words, data processing and distinguishing approaches are summarized and reviewed in these contributions. Channel selection, time window setting, and artifacts removal are employed in data preprocessing component. For feature extraction component, motor imagery (MI)-based EEG, steady state visual evoked potentials (SSVEPs), steady-state somatosensory evoked potentials (SSSEPs), and P300 are applied. Finally, features can be distinguished by linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbors (k-NN), and artificial neural networks (ANNs). The current review evaluates EEG-based BCI paradigms regarding their advantages and disadvantages from a variety of perspectives. For each paradigm, various EEG decoding algorithms and classification methods are evaluated and their application [19].

A new approach is raised according 'Think it aloud' dataset for EEG differentiation in this contribution. In the first step, electrodes that relevant to inner speech in headcap are chosen. Afterwards, selected data are decomposed by empirical mode decomposition (EMD) and intrinsic mode functions (IMFs) are acquired. Then support vector machine (SVM) is put into use for IMFs classification. To evaluate the performance of trained models, results are shown in two schemes: individual scheme and all subjects scheme. After suitable IMFs and kernels in SVM are employed, all classification values from both scheme are very good when EEG data are distinguished by the proposed approach.

Literature review

From the second half of the 20th century, inner speech has been a research topic in philosophy and psychology [20]. Varieties of Inner Speech Questionnaire (VISQ) has been used to link everyday phenomenology of inner speech – such as inner dialogue – to various psychopathological traits [21]. Apart from questionnaires, a large number of experiments are conducted to evaluate the effects of IS on intellectual development and psychiatric disorders. Eight chronic post-stroke aphasic patients and thirteen cognitively healthy adults are underwent testing on a range of evaluative tests and four experiments designed to check whether chronic post-stroke patients has the ability to use inner speech [22]. Referring to the effect of overt speech on children's use of inner speech in short-term memory, three experiments are implemented in [23], meanwhile, the role of private speech and inner speech in planning during middle childhood is tested in [24]. Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills is tested in experiments [25]. Direct evidence that inner speech sustains predictable task switching in adults are validated in [26]. University students are asked to listen to instrumental music and refer inner thoughts in a retrospective video-assisted interview to explore functions of inner speech and its expression in gesture [27]. Neural correlates focus of inner speech and auditory verbal hallucinations experiments are examined by Jones et al. [28]. Two experiments are designed to compare electrophysiological signals between inner speech and overt speech [29]. Lexical bias and the phonemic similarity effect in inner speech is validated in [30, 31]. To investigate the qualitative influence of inner speech on high and low measures of executive functions, two experiments are conducted [32]. Three experiments are carried out to examine the role of inner speech in task switching [33]. Usually, experiments related to IS are conducted by measuring signals from expressing of phonemes, words, vowels, and phrases.

Patterns capturing brain waves relevant to inner speech can be divided into invasive and non-invasive. Since invasive data acquisition requires surgery operation, typically, this pattern is not popular for human. However, Stephanie et al. [5] acquire ECoG signals regarding inner speech from electrode grids, strips or depth electrodes that are temporarily implanted onto the cortical surface, either above or below the dura mater. Comparing with invasive pattern, non-invasive pattern is more friendly for subjects who attend experiments. Through non-invasive experiments, MEG, fMRI, fNIRS, EEG, and sEEG signals are accessible. Magnetoencephalography (MEG) signals with high magnetometers and gradiometers on inner speech are obtained by Dash et al. [34]. Neural activity during inner speech of meaningless syllable sequences was measured with MEG and fMIR from eight right-handed subject in [35]. Functional near infrared spectroscopy make use of electromagnetic rediation in near-infrared region in order to measure functional activation in cortical areas 1-3 cm beneath the scalp. Kamavuako et al. obtain fNIRS signals referring to inner speech [36]. Comparing with ECoG, MEG, fMRI, and fNIRS signals, EEG signals are most widely used for inner speech analysis. Experiments gotten EEG data concerning inner speech data are conducted by [37, 38, 39, 40, 41, 42, 43, 44, 45]. Both fNIRS and EEG data are obtained within eleven participants performing multiple iterations of three tasks by Rezazadeh et al. [46]. Both EEG and fMRI data are recorded from ten healthy participants during covert speech in [47]. In addition, Angrick et al. record intracanial neural activity during speech processes using stereotactic electroencephalography (sEEG) electrodes [48].

Vowels, syllables, words, sentences, and states are performed by subjects and corresponding brain wave signals are acquired. To distinguish these signs, various traditional machine learning and deep learning approaches are applied. As the most classical machine learning method, support vector machine (SVM) is utilized by [36, 42, 37, 39, 43] to distinguish inner speech data. Nguyen et al. [40] use the variant of relevance vector machine (RVM) - a variant of SVM - to distinguish signs in experiments. Random forest (RF) which establishes outcome based on the decision trees predictions is utilized in [42, 39, 43] for inner speech distinction. As a classification and dimensionality reduction technique, linear discriminant analysis (LDA) is used in [46, 48, 42, 44]. Because of its simplicity and efficiency, Naive Bayes (NB) is also applied for inner speech distinction [43]. K-Nearest Neighbor (k-NN) which is a non-parametric supervised learning classifier is used for EEG data separation [45]. Another traditional machine learning method applied on fMRI and EEG data distinction is sparse logistic regression (SLR) [47]. Among deep learning, bidirectional long short-term memory recurrent neural network (BLSTM-RNN) which combine BLSTM and RNN are applied in [34]. Convolutional neural networks (CNN) is powerful for image distinction, is used for inner speech classification by [42, 38]. Deep Belief Networks (DBNs) invented as a solution for the problems encountered when using traditional neural networks training in deep layered networks is applied for inner speech discriminating [37].

The performance of approaches designing in some contributions of aforementioned should be accessed. The metrics for evaluating the performance of these approaches and classifiers are usually accuracy (ACC), standard deviation (SD), and F-score. Accuracy is most applied in literature. Besides, standard deviation is also employed to evaluate the results inequality among different subjects.For the imbalance classes, F-score can access the trained models unprejudiced. Results from part of the contributions mentioned above are shown in Table 1.

Table 1 Overview of inner speech studies

Data type	Subject number	Number of classes	Classifier	Results	Reference
fNIRS	8	6 words	SVM	ACC: 86.81, SD: 9.9	[36]
fNIRS, EEG	11	3 states	RLDA	ACC: 70.45 % SD: 19.19	[46]
EEG	15	3 vowels, 5 words	RVM	ACC: 70.00 - 95.00 %	[40]
EEG	10	4 words	EEG Net	ACC: 29.67 %	[41]
EEG	15	5 vowels, 6 words	SVM	ACC: 18.71-22.25 %	[42]
EEG	15	5 vowels, 6 words	RF	ACC: 18.37-23.23 %	[42]
EEG	15	5 vowels, 6 words	RLDA	ACC: 20.77-26.66 %	[42]
EEG	15	5 vowels, 6 words	CNN	ACC: 24.35-29.58 %	[42]
EEG	15	5 vowels, 6 words	EEGNet	ACC: 24.46-30.25 %	[42]
EEG	12	7 syllabic, 4 words	SVM	ACC: 18.08-79.16 %	[37]
EEG	12	7 syllabic, 4 words	DBN	ACC: 80.00-91.00 %	[37]
EEG	15	5 vowels	CNN	ACC: 32.75-34.41; F-score: 33.17-36.65 %	[38]
EEG	15	5 vowels, 6 words	SVM	ACC: 18.26-21.94 %	[39]
EEG	15	5 vowels, 6 words	RF	ACC: 19.60-22.72 %	[39]
EEG	27	5 words	RF/NB/SVM	best ACC: 58.41 % SD: 12.41	[43]
EEG	4	2 words	K-NN	ACC: 58.00 %	[45]

Theoretical preliminary

Although a large number of approaches applied for inner speech signals classification are designed and studied, the results of most these approaches shown in Table 1 are improvable. A new approach that combines data selection, EMD, Hilbert transform, and SVM will therefore be introduced. Before introducing the approach, theoretical preliminary will be introduced here.

Hilbert-Huang transform

Hilbert-Huang transform is suitable on performing time-frequency analysis on non-stationary and nonlinear data [49]. Two main steps are concluded in HHT: empirical mode decomposition and Hilbert spectral analysis. In the first step, the original signal is decomposed into a finite set functions which is known as intrinsic mode functions (IMFs) through an iterative process. The steps are:

1) Determine the local extreme of the signal;

2) Connect the maxima and minima with and interpolation function, generating an upper and a lower envelope about the signal;

3) Calculate the local mean as half the difference between the upper and lower envelopes. Subtract the local mean from the signal;

4) Iterate calculations on the residual.

Iterative process is repeated until the signal meets the definition of an IMF which is defined as signal with zero-mean, and its number of extreme and zero-crossing differ by at most one.

Second step of HHI is transforming IMFs from time domain to time-frequency domain by Hilbert transform. According to [50], equation 1 is known for Hilbert transform

$$H[g(t)] = g(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{g(\tau)}{t-\tau} d_\tau = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{g(t-\tau)}{\tau} d_\tau, \tag{1}$$

with

$H[g(t)]$: Hilbert transform,

$g(t)$: original signal.

Support vector machine

Support vector machine is one of the most popular supervised learning algorithms primarily used for classification problems. Goal of SVM is to generate the best line or decision boundary that can segregate n-dimension space into classes so that new data can be easily sorted in the correct category [51]. The best decision boundary is named as hyperplane. Support vector machine chooses the extreme points/vectors which are called support vectors to help in generating the hyperplane.

Support vector machine can be categorized into two types: linear and non-linear [52]. Mapping from the input space into the feature space is explained as well as the 'kernel trick' in SVM. A kernel function can be interpreted as a kind of similarity measure between input objects [53]. Various kernels can be used such as: linear, polynomial, and Gaussian kernels. In the proposed approach, linear kernel, different order polynomial kernel, and varied Gaussian kernels are tried.

Linear kernel function:

$$k(x_i, x_j) = 1 + x_i^{'} x_j, \tag{2}$$

Polynomial kernel function:

$$k(x_i, x_j) = (1 + x_i^{'} x_j)^p, \text{and} \tag{3}$$

Gaussian kernel function as

$$k(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{2\delta^2}), \tag{4}$$

with

x_i, x_j : features or data points,

p : polynomial order,

δ : kernel scale, and

n : number of data points.

According to the δ calculation formula, Gaussian SVM can be divided into fine Gaussian, medium Gaussian and coarse Gaussian.

Metrics

Following [54], to achieve an optimized classifier, diverse metrics play a crucial role for evaluation of ML approaches. Evaluation metrics are utilized through the two main stages of the generation of a usual ML classifier: training and test. The optimized solution of the training stage is discriminated and selected using the evaluation metric. Therefore, evaluation metrics are utilized to measure the efficiency of the generated class-related model. For test, the selected classifier gets evaluated considering the evaluation metric. Then results for test stage and the efficiency on training stage are compared to show the classifier performance on untrained data.

To evaluate a trained model performance, the classified data have to be categorized into true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Numbers of TP and TN are the samples number that are correctly classified [54]. Numbers of FP and FN are the samples number that have been misclassified [54]. The most common metrics for classification are accuracy, recall, precision, and F-score. Accuracy calculates the ratio of correct predicted cases to the overall number of evaluated samples. Recall or Sensitivity calculates the ratio of TP over the total number of TP and FN. Precision calculates the ratio of TP over the total number of TP and FP . F-Score] calculates the harmonic average between recall and precision. Every metric has its specific pros and cons: accuracy assesses the overall effectiveness of algorithms, precision assesses the predictive power of algorithms, sensitivity and specificity access the effectiveness of the algorithm on a single class; F-score benefits algorithms with higher sensitivity and challenges algorithms with higher specificity.

To evaluate the proposed approach comprehensively, models are tested in two schemes: individual scheme and all subjects scheme. In individual model, both training and test data come from the same subject. In all subject scheme, training data come from all subjects and test data come from part of each subject. To measure the distribution of models trained by each subject, standard deviations (SD) is applied as

$$SD = \sqrt{\frac{|x - \hat{x}|^2}{n}}. \tag{5}$$

Methodology

The dataset applied in this study is 'Thinking out lout' conducted by Nieto et al. [55]. Numerous data are acquired in this dataset while 128 electrodes are located in the headcap. Therefore, the first step is data selection. Afterwards, features of selected data are extracted by HHT. Furthermore, these features are distinguished by SVM. The details of the proposed approach are introduced as follows.

'Thinking out loud' dataset

Owing to lack of publicly available electroencephalography datasets restricts the development of new techniques for inner speech recognition, Nieto et al. [55] conducted an experiment to provide an open-access multiclass electroencephalography database of inner speech commands to the scientific community. In the experiment, ten healthy right-handed subjects – four females and six males with mean age ± std = 34 ± 10 years – without any hearing or speech loss, participated in the experiment. Electroencephalography headcap with 128 electrode were placed on each subject before the experiment. Besides headcap electrode, two computers and one computer screen graphic user interface (GUI) are included in the test as shown in Figure 1.

Each subject participated in one single recording day comprising three consecutive sessions. Within each session, five stimulation runs were presented. Four trial classes (words) – 'Arriba (up)', 'Abajo (down)', 'Derecha (right)', and 'Izquierda (left)' – are selected and presented in the screen. The trial's class (word) is selected randomly. The occurrence frequency of each word is the same, 559 times for each

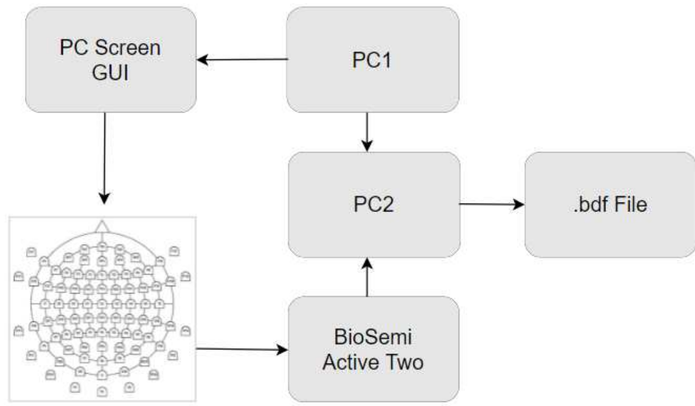

Figure 1 Test rig (inspired by [55])

word. In total, 2236 trials are held for inner speech. The detailed trial number of each subject is shown in Table 2.

Table 2 Numbers of trials of each subject [55]

Subjects	Up	Down	Right	Left
Sub-01	50	50	50	50
Sub-02	60	60	60	60
Sub-03	45	45	45	45
Sub-04	60	60	60	60
Sub-05	60	60	60	60
Sub-06	54	54	54	54
Sub-07	60	60	60	60
Sub-08	50	50	50	50
Sub-09	60	60	60	60
Sub-10	60	60	60	60
Sub total	559	559	559	559

According to [55], the procedure of each trial is as follows. Each trial begin at time $t = 0$ s with a concentration interval of 0.5 s. A white circle appears in the middle of the screen and participant should fix his/her gaze on it. Participants should not blink until the white circle disappear. At time $t = 0.5$ s the sign interval started. A white triangle pointing to different direction is presented. The pointing direction of the sign corresponds to each class. At $t = 1$ s, the triangle disappears from the screen and only the white circle remains. At the same time, the action interval started. Participants are instructed to start performing the indicated task right after the visual cues disappeared. After 2.5 s of action interval, i.e. at $t = 3.5$ s, the white circle turned blue and relax interval began. At $t = 4.5$ s the blue circle vanished and one trial ended. The workflow of each trial is shown in Figure 2.

Proposed approach

A large number of data are available in the dataset. However, not all data are needed as i) not all brain neurons are involved in the process of inner speech; ii) in case all data are in use, the calculation time is too long. Therefore, data selection is the first step to distinguish data. After data selection, EMD is applied to extract the main signal features. Then, features are classified by SVM. The flowchart of proposed approach is shown in Figure 3.

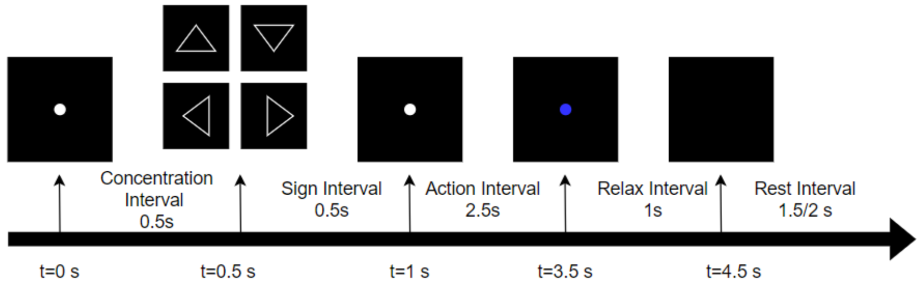

Figure 2 Trial workflow (inspired by [55])

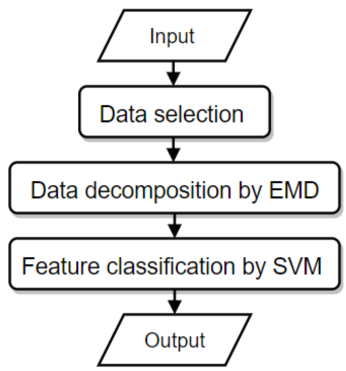

Figure 3 Flowchart of proposed approach

Data selection

When data from each electrode are used to differentiate the EEG signals, results are poor as it can be seen from the schematic results illustrated in Figure 4. The best F-score for a single electrode is less than 32 % [56]. It seems not to be necessary to calculate data form all electrodes because the speech-related neurons do not cover the whole cranial. Consequently, channel selection is necessary for data calculation.

According to [57], the ability to decode speech in a meaningful manner is complex and involves multiple stages of neural processing. Around 90 % of humans prefer their right hand for unimanual actions and are left-hemisphere dominant for language functions [58]. In the past, it was supposed that language was associated with the activity of three areas in the left hemisphere: the posterior frontal lobe, the upper segment of the temporal lobe, and the insula. Nowadays, most researchers assume that the cortical regions of the brain associated with the comprehension of language are Wernicke's area and the Broca's area [59]. In Wernicke's area, all language comprehension are controlled, in Broca's area all language production and the transmission of information between these areas is facilitated by the related arcuate fasciculus [60]. Besides, it is rational to consider that the primary motor cortex may also show similar hemispheric specialization for speech production [61]. Representation of continuous speech in the primary auditory cortex neurons is measured and how how individual phonemes modulate activity across the population of

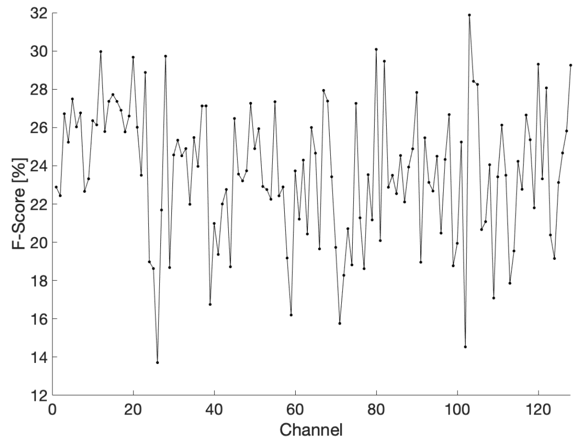

Figure 4 Results from each channel

auditory is examined in [62]. According to [63], the supramarginal gyrus is involved in phonological and articulatory processing of words, whereas the angular gyrus is involved in semantic processing. The combination of different parts in cortex in the human head is relevant to speech. Correspondingly, the related area is measured by channel D5 to D32 (electrodes inside of red circle) in the headcap as shown in Figure 5.

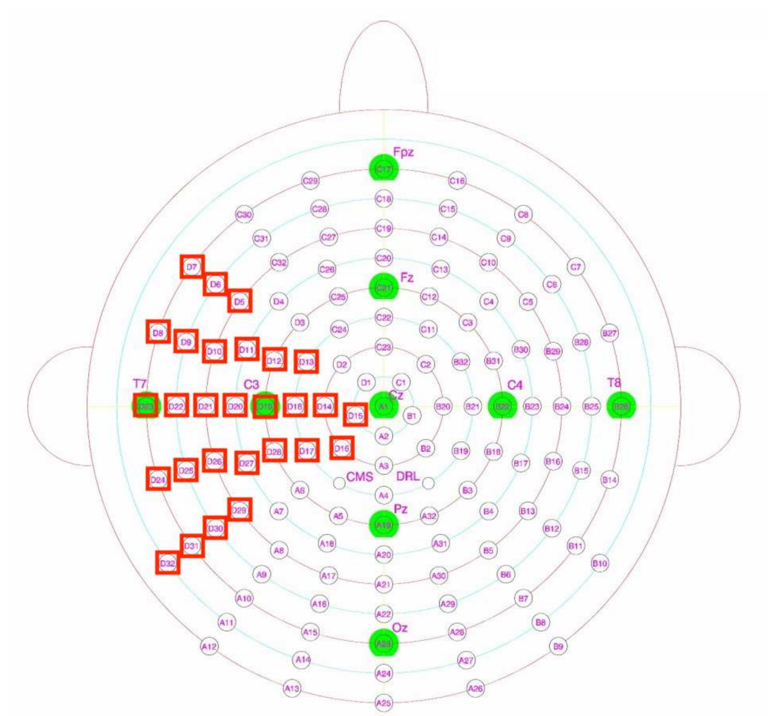

Figure 5 Electrodes selection

When data from channel D5 to D32 are used for calculation and linear SVM is applied for classification, results improved strongly as shown in Table 3. Average accuracy and F-score from individual scheme are over 93 %. Average accuracy and F-score from all subjects scheme are over 51 %. This states that channel selection has a significant influence on the classification results.

Table 3 Results from selected electrodes data

	Individual scheme		All subjects scheme	
Subjects	Accuracy	F-score	Accuracy	F-score
Sub-01	91.55	91.55	49.52	49.44
Sub-02	85.52	85.47	51.73	51.73
Sub-03	97.62	97.63	51.83	51.93
Sub-04	97.12	97.09	53.47	53.33
Sub-05	90.67	90.63	50.10	50.05
Sub-06	87.87	87.83	53.93	53.94
Sub-07	93.45	93.42	50.00	49.89
Sub-08	92.86	92.82	52.85	52.65
Sub-09	96.13	96.08	51.20	51.13
Sub-10	98.41	98.40	51.13	51.12
Average	93.12	93.09	51.58	51.52
SD	4.09	4.10	1.42	1.42

Data decomposition

Global brain activity is conventionally measured using electroencephalograms comprised in oscillations in several functionally-relevant frequency bands [64]. The bands are identified as δ (0.5-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-35 Hz), γ (35- Hz) waves [65]. According to [66], at the end of the arousal spectrum individual is basically disassociated from external world and exhibits a predominance of the δ band. With a predominance of the θ band, the individual focus is on the internal world. A predominance of β waves signal a state encompassing the thinking process with its accompanying ego reactions. The brain waves of α may be considered a bridge from the external world to the internal world. The brain wave band of γ is measured between (35-44 Hz) is the only frequency group found in every part of the brain. Based on the characteristics, selected measurements are decomposed into five IMFs. Raw data and its IMFs are shown in Figure 6. Transferring each IMF using Hilbert transform corresponding brain waves bands become visible, they are corresponding to the brain waves bands. That is: γ band results from spectrum of IMF1; β band results from spectrum of IMF2; α band results from spectrum of IMF3; θ band results from spectrum of IMF4; and δ band results from spectrum of IMF5. An example is IMF1 and the corresponding spectrum as shown in Figure 7.

To verify which brain waves bands represents the most relevant characteristics of inner speech, single IMFs and combinations are used for SVM-based classification. The F-score of individual subject schemes is shown in Table 4. From IMF1 to IMF4, results are getting worse, however, when IMFs are combined, results are improving especially the combination of IMF1, IMF2, and IMF3 generates the best results.

Feature classification

All previous calculations are based on the usage of linear kernel SVM. While different kinds of kernels could be applied in SVM, in this step various SVM kernels are evaluated applied on the same data set. In the setting of using selected data

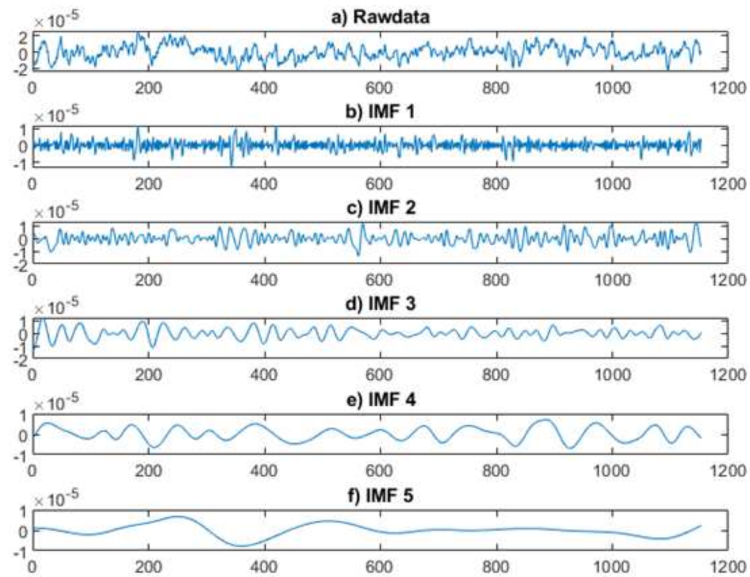

Figure 6 Raw data and IMFs: a) Raw data; b-f) Intrinsic mode functions

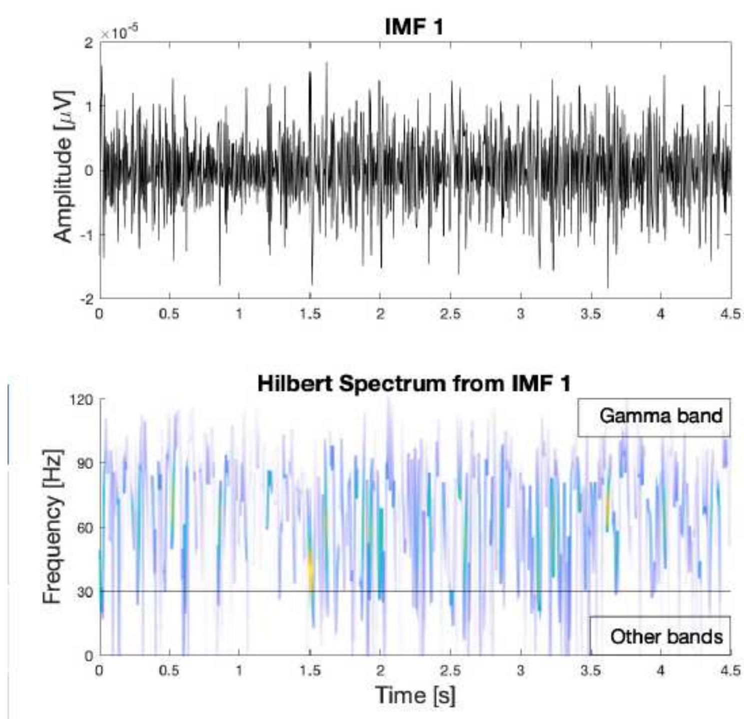

Figure 7 IMF and corresponding spectrum

and individual scheme, accuracy values are shown in Table 5. Except for the coarse Gaussian kernel, good results are obtained using other kernels to differentiate EEG

Table 4 Results (F-score) from different IMFs

	Single IMF				Combination of IMFs		
Subjects	1st	2nd	3rd	4th	1st and 2nd	1st to 3rd	1st to 4th
Sub-01	92.02	73.62	50.61	42.61	94.14	94.03	93.09
Sub-02	88.77	65.86	47.15	36.83	88.62	88.99	89.29
Sub-03	97.07	84.82	53.52	45.40	99.09	99.22	99.48
Sub-04	98.82	82.86	58.06	45.26	98.54	98.45	97.73
Sub-05	90.36	74.01	54.77	38.78	91.09	92.37	92.19
Sub-06	89.44	71.88	54.80	40.09	91.76	93.40	92.52
Sub-07	87.88	63.49	45.17	36.20	91.44	93.19	94.40
Sub-08	96.41	82.16	61.43	42.34	96.27	96.88	97.49
Sub-09	95.21	84.64	57.24	45.22	96.28	97.51	96.62
Sub-10	97.40	81.23	57.39	46.13	98.21	99.02	98.63
Average	93.34	76.46	54.01	41.89	94.54	95.13	95.04
SD	3.87	7.41	4.81	3.53	3.47	3.23	3.20

data. However, both accuracy and standard deviation from the 4th polynomial kernel are the best.

Table 5 Results from different SVM kernels

	Linear kernel	Polynomial kernel			Gaussian kernel		
Subjects	Linear	2nd	3rd	4th	Coarse	Medium	Fine
Sub-01	91.55	97.74	98.57	97.86	73.93	98.45	92.90
Sub-02	85.52	92.16	93.25	93.95	64.68	92.26	78.47
Sub-03	97.62	99.34	99.74	100.00	75.40	99.87	98.81
Sub-04	97.12	99.80	99.60	99.60	78.47	99.50	94.51
Sub-05	90.67	97.12	97.82	99.92	78.37	96.03	94.35
Sub-06	87.87	95.59	94.49	94.38	66.92	93.61	66.92
Sub-07	93.45	98.71	99.01	99.31	74.21	98.21	93.25
Sub-08	92.86	98.57	98.69	98.45	74.05	97.98	96.07
Sub-09	96.13	99.80	98.71	99.40	79.66	99.21	97.82
Sub-10	98.41	99.70	99.40	99.11	87.30	98.81	94.44
Average	93.12	97.85	97.93	97.90	75.30	97.39	90.63
SD	4.09	2.29	2.12	2.06	6.08	2.46	9.53

Best results and discussion

The evaluation of trained models is accessed. The metrics applied in this study are accuracy, F-score, and standard deviation. Accuracy assess the overall effectiveness while F-score gives better measure of the incorrectly classified cases. Furthermore, to access the discrepancy among each subject, standard deviation is also applied.

As data are decomposed into various IMFs in the process of EMD, each IMF and related combinations of them are evaluated. In addition, different kernels can be applied in features classification. Therefore, a large number of combinations is calculated. Best results are obtained when features are from combination of IMF1, IME2, and IMF3 in data decomposition step with 4th polynomial kernel in SVM. Detailed results (best results) are shown in Table 6.

The following conclusion can be drawn from Table 6:

i) All results show values higher than 95 % which denotes very high classification quality.

ii) In individual schemes, both accuracy and F-score for subjects 02 and subjects 03 show results of 100 % denoting that the models trained by single subjects could distinguish other data from the same subjects totally.

iii) Both accuracy and F-score for subject 8 are 100 %, this denotes that the trained model from all subjects schemes can distinguish subject 8 data totally.

Table 6 Best results

	Individual scheme		All subject scheme	
Subjects	Accuracy	F-score	Accuracy	F-score
Sub-01	99.64	99.63	99.11	99.09
Sub-02	96.83	96.83	99.85	99.86
Sub-03	100.00	100.00	99.81	99.80
Sub-04	100.00	100.00	99.71	99.71
Sub-05	98.61	98.61	99.58	99.57
Sub-06	98.57	98.57	99.83	99.83
Sub-07	99.90	99.91	99.53	99.54
Sub-08	99.29	99.27	100.00	100.00
Sub-09	99.80	99.81	99.85	99.84
Sub-10	99.80	99.80	99.40	99.40
Average	99.24	99.24	99.67	99.66
SD	0.95	0.95	0.27	0.27

iv) Results discrepancy between individual schemes and all subjects schemes are low.

v) A slight difference between accuracy and F-score as samples number in each class is observed.

vi) Standard deviation among different subjects is small in both individual schemes and all subjects schemes.

vii) The results from the proposed approach are much better than results from literature [36, 46, 40, 41, 42, 37, 38, 39, 43, 45]. This states that the proposed approach outperforms other approaches.

viii) Compared with [41] using the same dataset, the obtained results are significantly improved.

From the results in data selection, decomposition, and classification steps, the following conclusions can be drawn:

i) The inner speech-related districts in headcap can be located. Cortes region covered electrodes from D5 to D32 are strongly associated with inner speech.

ii) Results from the combination of IMF1, IMF2, and IMF3 are better than results from single IMF or other IMF combinations. This denotes that features from these IMFs are more suitable for data representation. When IMF1, IMF2, and IMF3 are transformed by Hilbert transform, related brain wave bands are γ, β, and α. Therefore, the conclusion that γ, β, and α brain wave bands are motivated in the process of inner speech can be drawn.

iii) Support vector machine with polynomial kernel is more suitable for EEG data differentiation.

To verify the generalization of proposed approach, the proposed approach which combines data selection, EMD and SVM could be applied to other EEG datasets or used for other neural signals. Apart from the application of the proposed approach to other datasets, other machine learning approaches such as convolution neural networks or generative adversarial networks can also be employed.

Acknowledgements
We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen

Funding
Not applicable.

Availability of data and materials
The dataset used for this research work is publicly available at
$https://github.com/N-Nieto/Inner_Speech_Dataset.$

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Wei carried out the background study, proposed and design the outline of proposed approach. Besides, she also write the first draft of this contribution. Surjana implement the calculation, take part in the design of proposed approach. Söffker initiated, raised, and supervised this study as well as revised and reviewed the text version of the paper.

Author details
Chair of Dynamics and Control, University of Duisburg-Essen, Duisburg, Germany.

References
1. Alderson-Day, B., Fernyhough, C.: Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychological bulletin **141**(5), 931 (2015)
2. Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., Loevenbruck, H.: What is that little voice inside my head? inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behavioural brain research **261**, 220–239 (2014)
3. Eisenberg, N., Sadovsky, A., Spinrad, T.L.: Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. New directions for child and adolescent development **2005**(109), 109–118 (2005)
4. Jani, M.P., Gore, G.B.: Occurrence of communication and swallowing problems in neurological disorders: analysis of forty patients. NeuroRehabilitation **35**(4), 719–727 (2014)
5. Martin, S., Iturrate, I., Millán, J.d.R., Knight, R.T., Pasley, B.N.: Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis. Frontiers in neuroscience **12**, 422 (2018)
6. Panachakel, J.T., Ramakrishnan, A.G.: Decoding covert speech from eeg-a comprehensive review. Frontiers in Neuroscience, 392 (2021)
7. Cheng, M., Gao, X., Gao, S., Xu, D.: Design and implementation of a brain-computer interface with high transfer rates. IEEE transactions on biomedical engineering **49**(10), 1181–1186 (2002)
8. Abdulkader, S.N., Atia, A., Mostafa, M.-S.M.: Brain computer interfacing: Applications and challenges. Egyptian Informatics Journal **16**(2), 213–230 (2015)
9. Abhang, P.A., Gawali, B.W., Mehrotra, S.C.: Technological basics of eeg recording and operation of apparatus. Introduction to EEG-and Speech-Based Emotion Recognition, 19–50 (2016)
10. Mahoney, V.M., Mezzano, V., Mirams, G.R., Maass, K., Li, Z., Cerrone, M., Vasquez, C., Bapat, A., Delmar, M., Morley, G.E.: Connexin43 contributes to electrotonic conduction across scar tissue in the intact heart. Scientific reports **6**(1), 1–12 (2016)
11. Burns, A., Adeli, H., Buford, J.A.: Brain–computer interface after nervous system injury. The Neuroscientist **20**(6), 639–651 (2014)
12. Huppert, T., Barker, J., Schmidt, B., Walls, S., Ghuman, A.: Comparison of group-level, source localized activity for simultaneous functional near-infrared spectroscopy-magnetoencephalography and simultaneous fnirs-fmri during parametric median nerve stimulation. Neurophotonics **4**(1), 015001 (2017)
13. Teplan, M., *et al.*: Fundamentals of eeg measurement. Measurement science review **2**(2), 1–11 (2002)
14. Zhang, W., Tan, C., Sun, F., Wu, H., Zhang, B.: A review of eeg-based brain-computer interface systems design. Brain Science Advances **4**(2), 156–167 (2018)
15. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for eeg-based brain–computer interfaces. Journal of neural engineering **4**(2), 1 (2007)
16. Al-Nafjan, A., Hosny, M., Al-Ohali, Y., Al-Wabil, A.: Review and classification of emotion recognition based on eeg brain-computer interface system research: a systematic review. Applied Sciences **7**(12), 1239 (2017)
17. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. Journal of neural engineering **15**(3), 031005 (2018)
18. Hwang, H.-J., Kim, S., Choi, S., Im, C.-H.: Eeg-based brain-computer interfaces: a thorough literature survey. International Journal of Human-Computer Interaction **29**(12), 814–826 (2013)
19. Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., Zhao, X.: A comprehensive review of eeg-based brain–computer interface paradigms. Journal of neural engineering **16**(1), 011001 (2019)
20. Vicente, A., Martinez Manrique, F.: Inner speech: Nature and functions. Philosophy compass **6**(3), 209–219 (2011)
21. Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., Fernyhough, C.: The varieties of inner speech questionnaire–revised (visq-r): replicating and refining links between inner speech and psychopathology. Consciousness and cognition **65**, 48–58 (2018)
22. Kljajevic, V., Gómez, E.U., López, C., Bandeira, Y.B., Vicente, A.: Inner speech in post-stroke motor aphasia. In: 39th Annual Meeting of the Cognitive Science Society (CogSci), pp. 2432–2437 (2017)
23. Hitch, G.J., Halliday, M.S., Schaafstal, A.M., Heffernan, T.M.: Speech,"inner speech," and the development of short-term memory: Effects of picture-labeling on recall. Journal of Experimental Child Psychology **51**(2), 220–234 (1991)
24. Lidstone, J.S., Meins, E., Fernyhough, C.: The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. Journal of Experimental Child Psychology **107**(4), 438–451 (2010)

25. Lidstone, J.S., Fernyhough, C., Meins, E., Whitehouse, A.J.: Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. Journal of autism and developmental disorders **39**(8), 1222–1225 (2009)
26. Laurent, L., Millot, J.-L., Andrieu, P., Camos, V., Floccia, C., Mathy, F.: Inner speech sustains predictable task switching: direct evidence in adults. Journal of Cognitive Psychology **28**(5), 585–592 (2016)
27. Fossa, P.: The representational and the expressive: Two functions of the inner speech. Psicologia: Teoria e Pesquisa **35** (2020)
28. Jones, S.R., Fernyhough, C.: Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. Clinical psychology review **27**(2), 140–154 (2007)
29. Whitford, T.J., Jack, B.N., Pearson, D., Griffiths, O., Luque, D., Harris, A.W., Spencer, K.M., Le Pelley, M.E.: Neurophysiological evidence of efference copies to inner speech. Elife **6**, 28197 (2017)
30. Brocklehurst, P., Corley, M.: Lexical bias and the phonemic similarity effect in inner speech. In: 15th Annual Conference on Architectures and Mechanisms for Language Processing, pp. 7–9 (2009)
31. Nooteboom, S.G.: Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. Speech communication **47**(1-2), 43–58 (2005)
32. Atkin, C.A., Badham, S.P.: The phenomenological influence of inner speech on executive functions. Journal of Undergraduate Research at NTU **1**(1), 223–251 (2018)
33. Emerson, M.J., Miyake, A.: The role of inner speech in task switching: A dual-task investigation. Journal of Memory and Language **48**(1), 148–168 (2003)
34. Dash, D., Ferrari, P., Berstis, K., Wang, J.: Imagined, intended, and spoken speech envelope synthesis from neuromagnetic signals. In: International Conference on Speech and Computer, pp. 134–145 (2021)
35. Fujimaki, N., Hayakawa, T., Matani, A., Okabe, Y.: Right-lateralized neural activity during inner speech repeated by cues. Neuroreport **15**(15), 2341–2345 (2004)
36. Kamavuako, E.N., Sheikh, U.A., Gilani, S.O., Jamil, M., Niazi, I.K.: Classification of overt and covert speech for near-infrared spectroscopy-based brain computer interface. Sensors **18**(9), 2989 (2018)
37. Zhao, S., Rudzicz, F.: Classifying phonological categories in imagined and articulated speech. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 992–996 (2015)
38. Cooney, C., Folli, R., Coyle, D.: Optimizing layers improves cnn generalization and transfer learning for imagined speech decoding from eeg. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 1311–1316 (2019)
39. Coretto, G.A.P., Gareis, I.E., Rufiner, H.L.: Open access database of eeg signals recorded during imagined speech. In: 12th International Symposium on Medical Information Processing and Analysis, vol. 10160, p. 1016002 (2017)
40. Nguyen, C.H., Karavas, G.K., Artemiadis, P.: Inferring imagined speech using eeg signals: a new approach using riemannian manifold features. Journal of neural engineering **15**(1), 016002 (2017)
41. van den Berg, B., van Donkelaar, S., Alimardani, M.: Inner speech classification using eeg signals: A deep learning approach. In: 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), pp. 1–4 (2021)
42. Cooney, C., Korik, A., Folli, R., Coyle, D.: Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech eeg. Sensors **20**(16), 4629 (2020)
43. González-Castañeda, E.F., Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L.: Sonification and textification: Proposing methods for classifying unspoken words from eeg signals. Biomedical Signal Processing and Control **37**, 82–91 (2017)
44. Song, Y., Sepulveda, F.: Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state. In: 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings, pp. 568–571 (2014)
45. Hashim, N., Ali, A., Mohd-Isa, W.N.: Word-based classification of imagined speech using eeg. In: International Conference on Computational Science and Technology, pp. 195–204 (2017)
46. Rezazadeh Sereshkeh, A., Yousefi, R., Wong, A.T., Rudzicz, F., Chau, T.: Development of a ternary hybrid fnirs-eeg brain–computer interface based on imagined speech. Brain-Computer Interfaces **6**(4), 128–140 (2019)
47. Yoshimura, N., Nishimoto, A., Belkacem, A.N., Shin, D., Kambara, H., Hanakawa, T., Koike, Y.: Decoding of covert vowel articulation using electroencephalography cortical currents. Frontiers in neuroscience **10**, 175 (2016)
48. Angrick, M., Ottenhoff, M.C., Diener, L., Ivucic, D., Ivucic, G., Goulis, S., Saal, J., Colon, A.J., Wagner, L., Krusienski, D.J., *et al.*: Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Communications biology **4**(1), 1–10 (2021)
49. Barnhart, B.L.: The hilbert-huang transform: theory, applications, development. PhD thesis, The University of Iowa (2011)
50. Johansson, M.: The hilbert transform. Mathematics Master's Thesis. Växjö University, Suecia. Disponible en internet: http://w3. msi. vxu. se/exarb/mj_ex. pdf, consultado el **19** (1999)
51. Tan, Y., Wang, J.: A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension. IEEE Transactions on knowledge and data engineering **16**(4), 385–395 (2004)
52. Ghosh, S., Dasgupta, A., Swetapadma, A.: A study on support vector machine based linear and non-linear pattern classification. In: 2019 International Conference on Intelligent Sustainable Systems (ICISS), pp. 24–28 (2019)
53. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, pp. 1458–1465 (2005)
54. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L.: Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. Journal of Big Data **8**(1), 1–74 (2021)
55. Nieto, N., Peterson, V., Rufiner, H.L., Kamienkowski, J.E., Spies, R.: Thinking out loud, an open-access

eeg-based bci dataset for inner speech recognition. Scientific Data **9**(1), 1–17 (2022)
56. Surjana, A.I.: Application of machine learning on brain speech recognition. Master's thesis, University of Duisburg-Essen (2021)
57. Chandrasekaran, B., Kraus, N.: The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology **47**(2), 236–246 (2010)
58. Papadatou-Pastou, M.: Handedness and language lateralization: Why are we right-handed and left-brained. Hellenic Journal of Psychology **8**(2), 248–265 (2011)
59. Dronkers, N.F., Wilkins, D.P., Van Valin Jr, R.D., Redfern, B.B., Jaeger, J.J.: Lesion analysis of the brain areas involved in language comprehension. Cognition **92**(1-2), 145–177 (2004)
60. Dronkers, N.F.: The pursuit of brain–language relationships. Brain and Language **71**(1), 59–61 (2000)
61. Terumitsu, M., Fujii, Y., Suzuki, K., Kwee, I.L., Nakada, T.: Human primary motor cortex shows hemispheric specialization for speech. Neuroreport **17**(11), 1091–1095 (2006)
62. Mesgarani, N., David, S., Shamma, S.: Representation of phonemes in primary auditory cortex: how the brain analyzes speech. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, vol. 4, p. 765 (2007)
63. Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., Frackowiak, R.: The anatomy of phonological and semantic processing in normal subjects. Brain **115**(6), 1753–1768 (1992)
64. Moffett, S.X., O'Malley, S.M., Man, S., Hong, D., Martin, J.V.: Dynamics of high frequency brain activity. Scientific reports **7**(1), 1–5 (2017)
65. Groppe, D.M., Bickel, S., Keller, C.J., Jain, S.K., Hwang, S.T., Harden, C., Mehta, A.D.: Dominant frequencies of resting human brain activity as measured by the electrocorticogram. Neuroimage **79**, 223–233 (2013)
66. Khalily, M.T., Clarke, L., Jahangir, F.: Alpha-theta neurofeedback therapy: A non-invasive treatment for addiction and post-traumatic stress disorder. FWU Journal of Social Sciences **3**(2), 83 (2009)