
School of Computer Science and Electronic Engineering

# A novel EEG based linguistic BCI

Thesis submitted for the degree of Doctor of Philosophy

By

**Amir Jahangiri**

Supervisor

**Francisco Sepulveda**

September 2021

## *Abstract*

While a human being can think coherently, physical limitations no matter how severe, should never become disabling. Thinking and cognition are performed and expressed through language, which is the most natural form of human communication. The use of covert speech tasks for BCIs has been successfully achieved for invasive and non-invasive systems. In this work, by incorporating the most recent discoveries on the spatial, temporal, and spectral signatures of word production, a novel system is designed, which is custom-build for linguistic tasks. Other than paying attention and waiting for the onset cue, this BCI requires absolutely no cognitive effort from the user and operates using automatic linguistic functions of the brain in the first 312ms post onset, which is also completely out of the control of the user and immune from inconsistencies. With four classes, this online BCI achieves classification accuracy of 82.5%. Each word produces a signature as unique as its phonetic structure, and the number of covert speech tasks used in this work is limited by computational power. We demonstrated that this BCI can successfully use wireless dry-electrode EEG systems, which are becoming as capable as traditional laboratory-grade systems. This frees the potential user from the confounds of the lab, facilitating real-world application. Considering that the number of words used in daily life does not exceed 2000, the number of words used by this type of novel BCI may indeed reach this number in the future, with no need to change the current system design or experimental protocol. As a promising step towards non-invasive synthetic telepathy, this system has the potential to not only help those in desperate need, but to completely change the way we communicate with our computers in the future as covert speech is much easier than any form of manual communication and control.

# Dedication

As long as the mind works, no disability should ever be disabling.
This work is the result of the most fulfilling years of my life.
I humbly dedicate it to my fellow man. May it be built upon forever.

# *Table of Contents*

# CHAPTER ONE:
## INTRODUCTION

## *1.1. Consciousness, cognition, and language in the brain*

The Human body samples the physical world through sensory organs and causes changes in the physical world by controlled muscle contractions. Even the most minute conscious movements can be effectively used in human-machine interaction (i.e. the late Prof. Hawking using ocular-facial tracking). In some circumstances, however, conscious control of muscles may be severely limited or completely lost (i.e. patients with locked-in syndrome) isolating a conscious thinking human from the world. A Brain-Computer Interface (BCI) is a system, which determines the "conscious" intention of a person by measuring and analysing brain activity and generates appropriate command signals for communication or control. So far, BCIs have been successfully used as assistive medical tools for the most severe cases.

Before engaging the technical issues related to BCIs, it is worth asking the question: what is consciousness? In recent years, much advancement has been made in the fields of Neuroscience and Physiology, which provide a greater insight into the operation of the brain. However, the exact nature of our consciousness remains unknown and perhaps this is one of the greatest medical, technical, and philosophical questions of all time. One interesting theory which tries to explain consciousness is "'the Conscious Electromagnetic Information (CEMI) field theory" [1, 2]. This theory, the electromagnetic field generated by

the brain is considered to influence the sensitive voltage gated Neuron channels in the membrane of Neurons. The theory states that digital information in the Neurons is summed up from an electromagnetic field containing information. Therefore, consciousness is that segment of the field generated by the brain that is downloaded to the movement sections of the brain and by movement can influence the outside world [1]. In other words, in addition to the unique electrical map of the in the brain, the unique EM field plays an essential role in the function of the brain. This raises the possibility of downloading and recreating the mind, in essence artificial resurrection [3]. Another interesting theory, which tries to explain the nature of consciousness, is the Extended Everett's Concept (EEC), which ties Quantum Physics in to the operation of the mind [4]. It states: "Consciousness is the inherent ability of living beings to perceive alternative classical projections of the objectively quantum world separately from each other". It also progresses a step further and states: "Super-Consciousness, or intuition (existing in the state of meditation, trance, or dream) provides access to all classical alternatives and usage of the obtained information [4].

To this date, BCIs have used detectable signatures of user's conscious intent, such as autonomic responses to stimuli or imagined movements (Motor Imagery). These approaches have limitations such as requiring constant attention or gaze leading to user fatigue, or inconsistency in imagined movements for MI leading to poor performance. Insufficient understanding of the brain, sensor's

signal to noise ratio, and computational cost have impeded the use of higher cognitive functions such as speech, especially in non-invasive BCIs. In the past few years, major research has been conducted on the process of generating speech, and the neurological and psychological functions underlying this process. Research suggests that it is possible to extract significant and meaningful data from EEG measurement of the brain, during mental activity related to the generation of covert speech. There is strong evidence that shows the activation of the Broca's area during auditory speech perception [5]. The Broca's area has involvement in some other functions related to imitation of movement [6]. One of the main functions of the Broca's region is generation of speech by sending signals to the frontal cortex and the pre-motor area. These signals are processed and sent to the motor cortex. Studies have been conducted on the motor signals generated by the brain during natural speech and it has been shown that motor signals related to language can be detected wit high imaging resolution. Imagined speech has been used to transmit the alphabet by producing two syllables in one of three different rhythms (similar to Morse code) [7]. These studies suggest that there is considerable information in different frequency bands, and feature vectors within each band associated with imagined speech. A similar study in generation of imagined generation of two syllables (“ba” and “ku”) was successfully used to identify states with good success [8]. In the above work, the EEG data was pre-processed to reduce the effects of artefacts and noise. Auto regressive coefficients were extracted from each electrode's signal and

concatenated for user identification. EEG has been extensively used in the study of language generation and phoneme monitoring. As mentioned, the focus is on event related potentials [9, 10]. However, the generation of covert speech is relatively new. In the study of EEG special care must be taken to remove muscular related artefacts from the signal [10]. To the present day, a maximum of eleven covert speech tasks three covert speech classes have been classified using EEG [11], with the majority only classifying three [12], or two word-classes successfully [7, 13-22]. In some cases, where due to some underlying medical condition brain surgery is required (i.e. treatment for epilepsy), sensors can be placed directly on the brain (ECoG). This method provides a much clearer source of information compared to EEG as there is no signal degradation due to the scattering effects of the scull and skin tissue. The focus of a number of ECoG measurements have been on language production [23-25], and in some cases covert speech [24]. These studies affirm the conventional findings of EEG methods in the location of the language centres (Superior temporal Gyrus STG), and increased band power in the higher Gamma band, in addition to increased average evoked potentials during speech generation (AEGs) [23]. In addition to confirming measurable effects, the location of the brain areas involved in speech generation have been especially studied with ECoG sensors [25-27]. These investigations have shifted from mapping brain areas related to speech generation [28], to testing hypotheses of how the brain works as whole to generate language [27]. Based on all the findings related to speech generation, a model that closely

describes the experimental results is the internal forward model [29, 30]. In this model, the motor state is affected by estimated motor state, predicted perceptual consequence, and the somatosensory and perceptual feedback. The motor commands are planned in accordance to the intended movement. This signal is sent to: The peripheral motor system and a copy is sent to the first internal forward model for estimation of the following motor state. A second efference copy known as the perceptual efference copy is sent to the second internal forward model to predict the perceptual consequence of such motor estimation. Figure 1.1 demonstrates this model.

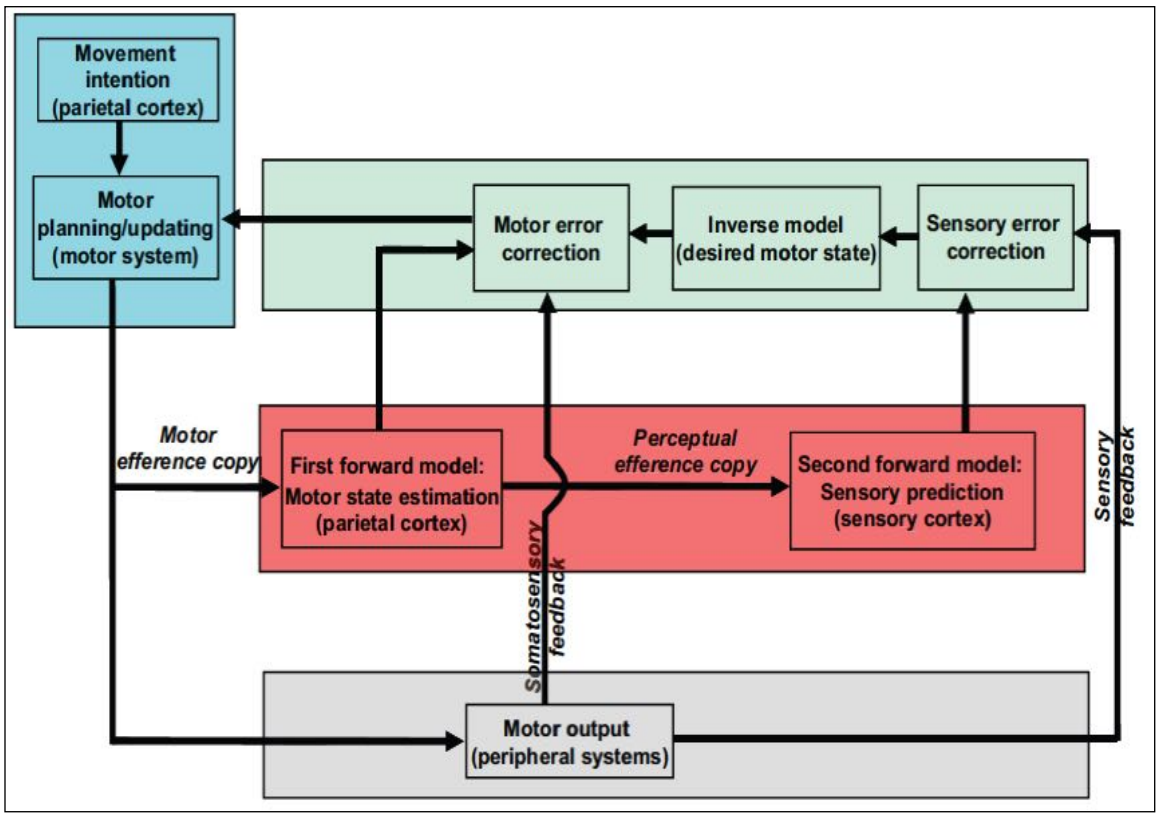



***Figure 1.1.*** *The model of motor control based on internal forward models and feedback [29].*

Although extensive research has been conducted on language, and the physiological and psychological functions related to it, not much has been done to use language for brain computer interfaces. A recent study in 2011 has successfully used imagined speech in conjunction with imagined limb movement to control a robotic wheelchair [31]. A recent study has been conducted to investigate the possibility of detecting speech using rhythmic activity of the brain [32]. With the use of Magneto Encephalography (MEG), the event related de-synchronization (ERD) related to vocalization or seeing a video of vocalization was measured. The measured MEG signals were then transformed into a time-frequency representation (TFR). Analysis of this data showed promise of distinguishing between tasks, with good accuracy. This method could possibly be used to create BCIs to help people with speech impairments. The presence of speech during imagination of movement has also been investigated [14]. This work suggests the possibility of identifying motor imagery with the presence (or in absence) of speech. The analysis of spectral event related power revealed the possibility to distinguish between movement, speech, and speech and movement.

## *1.2. Bioelectrical and imaging research on word production in the brain*

Now that we have briefly covered the fundamentals of language in the brain, it is time to focus on the specific linguistic activity we wish to use in a BCI. With regards to creating a BCI that focuses on using language, the linguistic

activity must be consciously generated by the user to communicate a intention. The most intuitive and natural way for this is generating words. There are four areas of interest. Firstly, what frequency range do the signatures of word production in the brain cover. Secondly, what is the exact sequence of activities in the brain leading to self-generated words. Thirdly, what is the exact temporal sequence and duration of these stages, and fourthly, where in the brain do these activities take place.

Much research has been done investigating covert speech production using ECoG measurements [33]. The best evidence for covert speech has been obtained from an invasive ECoG study by the USA military [34] using the general-purpose BCI2000 software connected to a “g.USBamp” amplifier and digitizer (g.tec, Graz, Austria). In this work the ECoG signals were amplified, bandpass filtered (0.15-500 Hz), digitized at 1200 Hz, and saved with BCI2000. Data was re-referenced using common average referencing. Every 10 ms, the time-series ECoG signals of the previous 167 ms were transformed into the frequency domain with an autoregressive model of order 25. With this model, the spectral amplitudes between 70 and 170 Hz in 2 Hz bins were discovered. With a window size of 167 ms, the temporal accuracy of the frequency estimates was ±83 ms. These spectral amplitudes within this band were averaged, generating a time course of high Gamma power for each electrode. To determine the task-related difference in ECoG high Gamma power for each electrode, the coefficient of determination (r^2) between the ECoG HG power for covert word production and

idle state were calculated. As a result, the location, time point, and the statistical difference between covert speech task and rest were discovered. In regard to the frequency signatures of word production, not only the Alpha and Beta bands are involved, but there are signatures in the Gamma, and high Gamma band up to 175Hz [27, 35-39], as seen in figure 1.2.

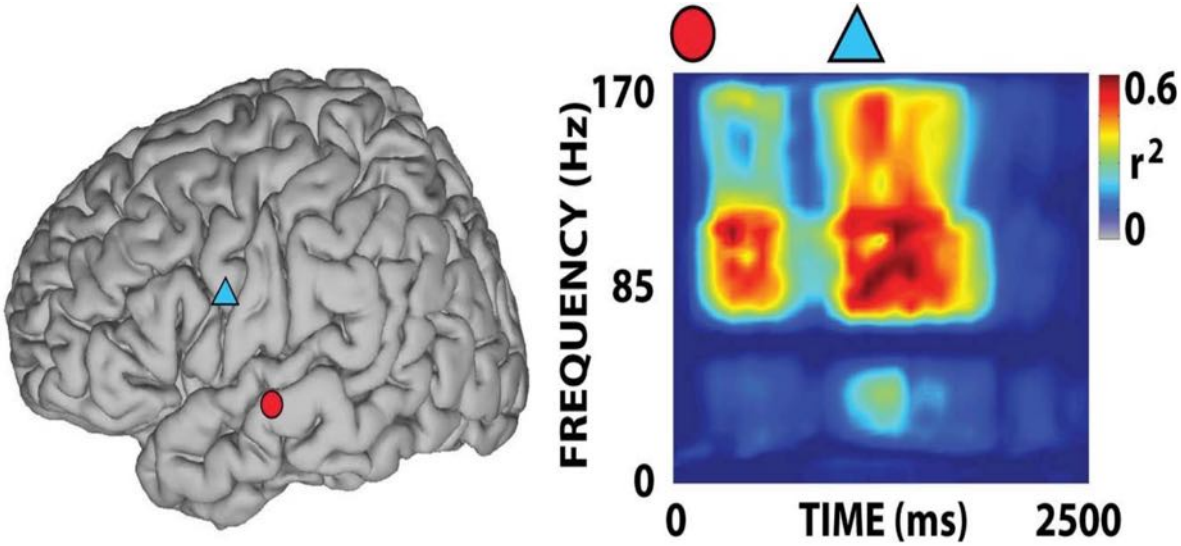



***Figure 1.2.*** *The frequency range and time course of neural activity of two points in the left hemisphere of the brain during covert word production* [34]

This is well above the traditional frequency range studied in non-invasive BCIs to this date. However, invasive studies have been critical in identifying not only the frequency of neural operations for word production, but also the sequence and timing of each brain region involved in this type of activity [25]. As seen in figure 1.3, these areas include Premotor area, primary motor area,

middle superior temporal gyrus, posterior superior temporal gyrus, Broca's Area, Angular gyrus, Wernicke's area, and the posterior middle temporal gyrus [34].

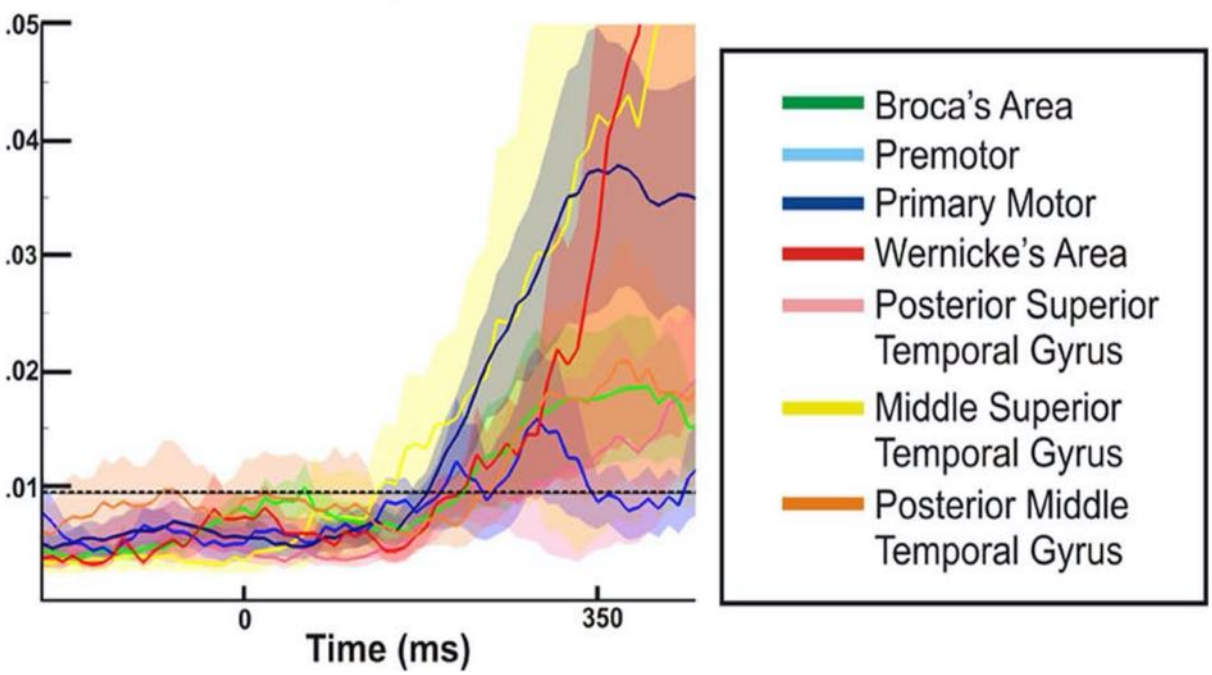



***Figure 1.3.*** *The activation of Gamma band in different areas of the brain involved with covert word production and the time of activation relative to auditory task onset [34]*

Word production begins with semantic (conceptual preparation), lexical (Lemma retrieval), and phonetic (phonological code retrieval and syllabification) linguistic processes, followed by planning the movements of language muscles (phonetic encoding) for articulation [27, 34]. Figure 1.4 illustrates the approximate time course of each stage of word production alongside the brain region involved in that stage [39].

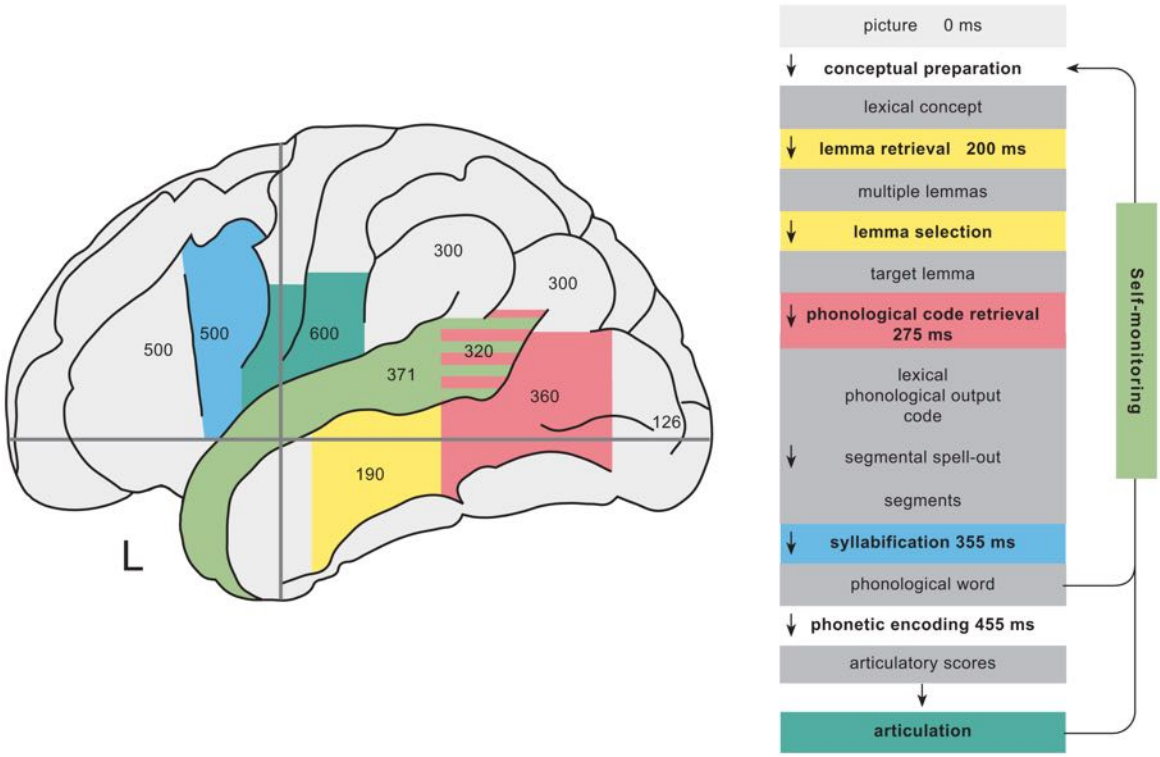



***Figure 1.4.*** *The staged of word production and their timing (right), and the respective brain areas for each stage (left). The focus here is on the left hemisphere [39].*

Linguistic phonetic processing is an automatic brain function, which elicits high-Gamma (70-170 Hz) oscillations [40, 41]. In each individual, Phonetic processing activity for a specific word does not change over time [42, 43] and is not affected by priming, cognitive activity, or task frequency [44, 45]. In contrast, semantic and lexical processing, is affected by task frequency, priming, and cognitive activity [9, 10, 25], which would also arbitrarily shift the temporal course of all following functions. These problems can be avoided by using a suitable experimental protocol. In covert speech, the manner of articulation in an individual (their 'accent') is consolidated over time. Covert articulation tasks activate the same language motor centres as their overt form [46, 47]. As a result,

covert speech is produced with the same consistency as overt speech. However, in covert speech, the activity of the Primary Motor Cortex is greatly attenuated [48] and may be difficult to detect by EEG.

## *1.3. Research Objectives*

Although the exact nature and operation of "consciousness" is not fully known, that aspect of consciousness called the intellect, which assists us in physical survival through cognition and thinking, is closely intertwined with language and Linguistic processes in the brain. Speaking is the most natural form of human communication and a BCI which can be controlled by thinking of words, would be extremely intuitive and easy to use. Considerable research has been conducted on language in the brain in the past decade, which have not yet been considered in the design of a new purpose-built linguistic BCI system.

In this work, I explore the capability of non-invasive EEG systems, both wet and dry electrode devices, in detecting activity in the Gamma and high Gamma bands and how to reduce their susceptibility to EMG and EOG artefacts. To this end the objective is to increase the traditional frequency range used in BCIs (usually up to 50Hz) and to take advantage in the improved signal to noise ratio of modern EEG equipment and push the frequency envelope as far as possible. This includes selecting the correct recording sampling rate and discovering the most capable artefact removal algorithms.

The next objective is to study the class separability of covert speech tasks compared to pure motor imagery tasks to determine how the combination of linguistic processing functions and speech motor functions in covert speech tasks differ from imagined movement for the purpose of classification.

After this, I aim to compare the contribution of pure linguistic processing stages or word production before articulation to the movement planning and motor imagery of articulation and determine which activities are most useful in class separation. With this knowledge, I finally aim to create an online BCI and refine each, and every stage of the analysis pipeline to maximise classification accuracy. The final goal is to demonstrate an online BCI based on the research in the previous chapters.

## *1.4. Hypothesis for the design of a Linguistic BCI*

***Hypothesis one:*** Imagined speech (covert speech) can be detected in the same way that imagined movement can be detected. This is by focusing on the motor cortex and taking measurements in different frequency bands (theta 3-8 Hz. Alpha 8-13 Hz, and beta 13-45 Hz) [49, 50]. Speech is the most complex movement generated by the brain and over half the Primary Motor Cortex is responsible for controlling muscles involved in speech production. The main muscles involved in speech production are shown in figure 1.5.

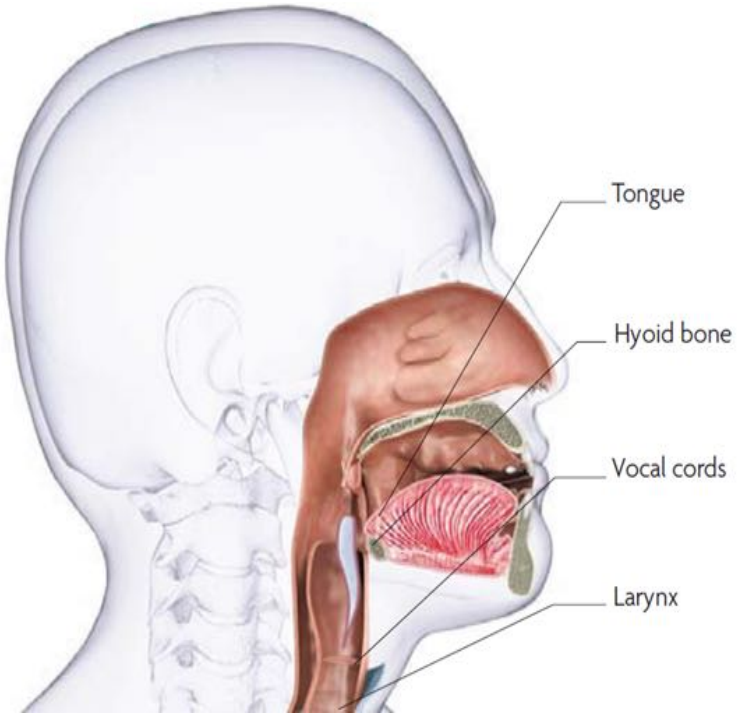



***Figure 1.5:*** *Muscles involved in the generation of speech [51].*

***Hypothesis two****:* Imagined speech can be detected by focusing measurements on the language centres of the brain (i.e. Broca's Area). Before the execution of speech by the relevant muscle groups (face, mouth, lips, Tongue, larynx, and lungs), there is a complex process of planning and linguistic processing. The order to execute speech is sent from the Broca's area to the frontal cortex and the pre-motor area and then to the motor cortex. This hypothesis suggests that it is possible to detect linguistic activity before the command to execute reaches the motor cortex.

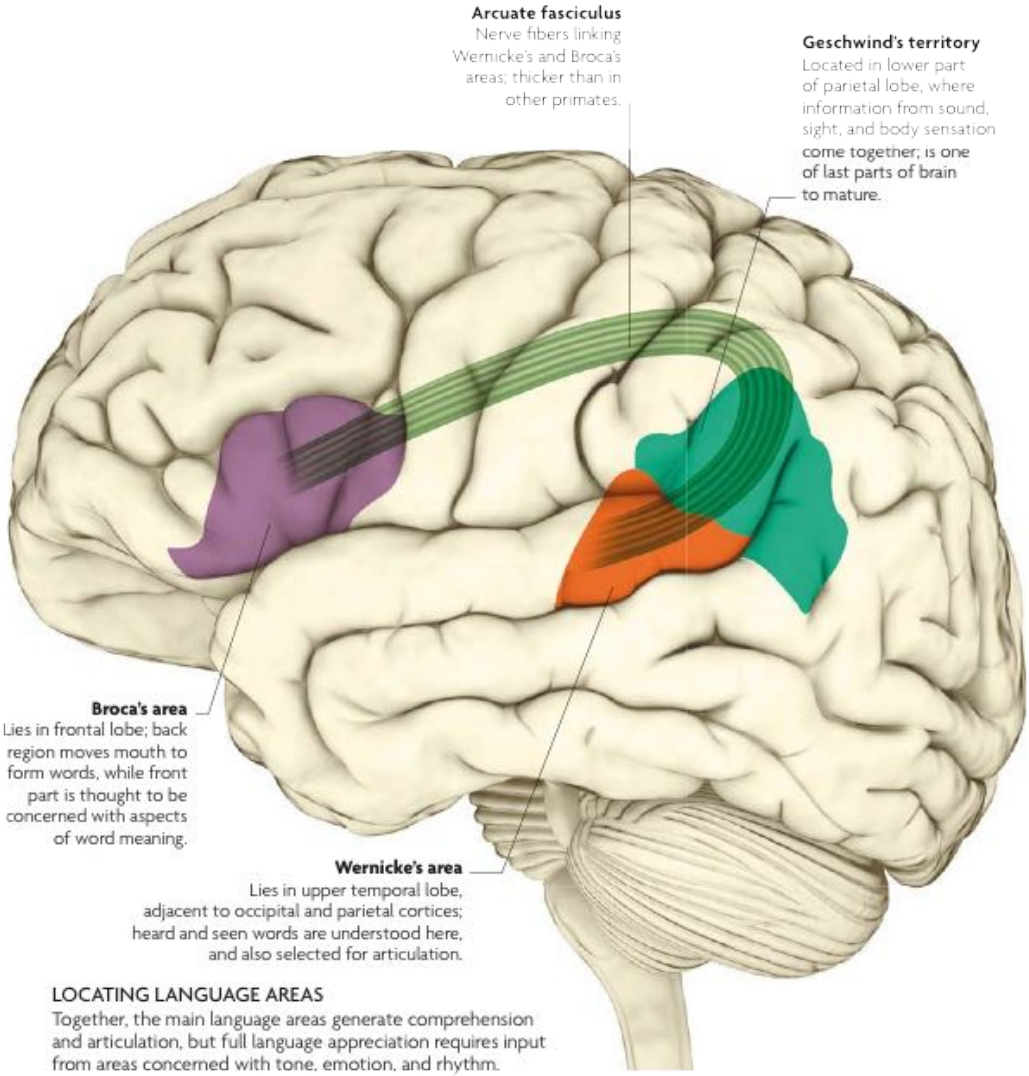



***Figure 1.6:*** *The Wernicke and Broca's area [51].*

***Hypothesis three****:* Before the involvement of the Broca's Area in the execution of speech (and the setting of prosody) there is a multitude of activities in the brain, which occur in known time intervals [9]. It could be possible to detect these event-related potentials in a similar wat to P300 systems. These Linguistic ERPs which can be used in a BCI include:

- N400, whole brain, negative peak, lexical-semantic processes
- ELAN, Left Anterior Negativity, Structure building

- P600, late ERP positive Wave, Syntactic integration
- The CPS component: Prosodic processing

***Hypothesis Four****:* Using language in a BCI requires a completely new approach by focusing on specific linguistic processing stages of word production, and using the respective temporal, spectral, and spatial signatures of a particular stage (for example phonological and phonetic stages). Such an approach has never been used.

## *1.5. Contributions*

The first publication from this work is titled: “The contribution of different frequency bands in class separability of covert speech tasks for BCIs” was published by IEEE EMB in 2017. In this work, ten volunteers in the age range of 22-70 years participated in the experiment. Eight of them were neurologically healthy, one user was dyslexic, and another was autistic. The four words “back”, “forward”, “left”, and “right” were shortened into “BA”, “FO”, “LE”, and “RY”, which are phonetically dissimilar and cognitively relevant directional commands. Participants were asked to covertly speak each as soon as the letters appeared on a screen. Volunteers completed five recording runs. During each run the four words were presented in random succession to avoid sequence bias. The recorded EEG data from the ten users were analysed to discover the best feature vectors within a Gabor Transform of the signals, i.e., those yielding the highest word-pair

classification accuracy for this specific type of linguistic mental activity. Using this BCI, suitable class separability of covert speech tasks is confirmed for all, including disabled users, with consistently high classification accuracy from 75% to 88% in all cases. Similar to motor imagery tasks, Alpha and Beta band activity were found to contain 12% and 31% of the most important feature vectors, respectively. Gamma band activity, which indicates high mental functions, contains 57% of the most important feature vectors in this study.

The second publication from this research with title “Covert Speech vs. Motor Imagery: a comparative study of class separability in identical environments” was published by IEEE EMB in 2018. In this study a single experimental protocol and analysis pipeline is used: once for MI tasks, and once for covert speech tasks. The goal of this study is not to maximize classification accuracy; rather the main objective is to provide an identical environment for both paradigms, while identifying the most important activities related to the most class dependent feature vectors. Four volunteers participated in this experiment. With four classes, the average classification accuracy for covert speech tasks is 82.5%, and for motor imagery is 77.2%. The average performance is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect. For motor imagery tasks the most important activities are the execution of imagined movements, and goal driven executive control for suppression of overt movements, which also occur for covert speech tasks. However, the most important activity for covert speech tasks is the

linguistic processing stages of word production prior to articulation, which does not occur in motor imagery. These high-Gamma linguistic processes are extremely class dependent, which contribute to the higher performance of covert speech tasks, compared to motor imagery in an otherwise identical environment.

The third publication from this work with title “The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data” was published by Springer’s journal of Medical Systems in 2019. Word production begins with high-Gamma automatic linguistic processing functions followed by speech motor planning and articulation. Phonetic properties are processed in both linguistic and motor stages of word production. Four phonetically dissimilar phonemic structures “BA”, “FO”, “LE”, and “RY” were chosen as covert speech tasks. Ten neurologically healthy volunteers with the age range of 21-33 participated in this experiment. Participants were asked to covertly speak a phonemic structure when they heard an auditory cue. EEG was recorded with 64 electrodes at 2048 samples/s. Initially, one-second trials were used, which contained linguistic and motor imagery activities. The four-class true positive rate was calculated. In the next stage, 312ms trials were used to exclude covert articulation from analysis. By eliminating the covert articulation stage, the four-class grand average classification accuracy dropped from 96.4% to 94.5%. The most valuable feature vectors emerge after Auditory cue recognition (~100ms

post onset), and within the 70-128 Hz frequency range. The most significant identified brain regions were the Prefrontal Cortex (linked to stimulus driven executive control), Wernicke's area (linked to Phonological code retrieval), the right IFG, and Broca's area (linked to syllabification). Alpha and Beta band oscillations associated with motor imagery do not contain enough information to fully reflect the complexity of speech movements. Over 90% of the most class-dependent feature vectors were in the 30-128Hz range, even during the covert articulation stage. As a result, compared to linguistic functions, the contribution of motor imagery of articulation in class separability of covert speech tasks from EEG data is negligible.

The fourth publication with title "A Novel EEG-Based Four-Class Linguistic BCI" was published in the EMBC conference in 2019. In this work, we present a novel EEG-based Linguistic BCI, which uses the four phonemic structures "BA", "FO", "LE", and "RY" as covert speech task classes. Six neurologically healthy volunteers with the age range of 19-37 participated in this experiment. Participants were asked to covertly speak a phonemic structure when they heard an auditory cue. EEG was recorded with 64 electrodes at 2048 samples/s. The duration of each trial is 312ms starting with the cue. The BCI was trained using a mixed randomized recording run containing 15 trials per class. The BCI is tested by playing a simple game of "Wack a mole" containing 5 trials per class presented in random order. The average classification accuracy for the 6 users is 82.5%. The most valuable feature vectors emerge after Auditory cue

recognition (~100ms post onset), and within the 70-128 Hz frequency range. The most significant identified brain regions were the Prefrontal Cortex (linked to stimulus driven executive control), Wernicke's area (linked to Phonological code retrieval), the right IFG, and Broca's area (linked to syllabification). In this work, we have only scratched the surface of using Linguistic tasks for BCIs and the potential for creating much more capable systems in the future using this approach exists.

## *1.6. Outline of thesis*

After the introduction in chapter one, the relevant background literature on brain computer interfaces and the use of language in BCIs are presented in the second chapter. Problems related to the curse of dimensionality and how they are addressed in this work are also in chapter two.

Chapter three discusses the selection of covert speech tasks and designing the first experimental protocol. This protocol was used to compare performance of the design for covert speech tasks vs Motor Imagery tasks in identical settings. The validity of the analysis pipeline was tested using random permutation tests, rest state vs. task tests, and using fewer features for classification.

In the fourth chapter, the experimental protocol was modified to increase the number of channels from 20 to 64 replacing the dry electrode system

with the Biosemi ActiveTwo system. In addition, the time duration of trials was reduced from one second, to 312ms with virtually no loss of performance. This new protocol is more suitable for online use. Before performing online experiments, the validity of the revised analysis pipeline was tested using random permutation tests, rest-state vs. task classification, using lees features for classification, of the topographical maps of the brain, and using HV-validation vs. 10-fold cross validation.

Chapter five presents the culmination of this research and presents an online demonstration of a Linguistic BCI using the experimental protocols and analysis pipeline designed in the previous chapter. The results in chapter five show that the system does indeed perform robustly (classification accuracy 82.5%) and that there was a 12.5% upward estimation bias in the results using offline data and block-designed validation processes.

# CHAPTER TWO:
## BCI SYSTEMS AND THE CURSE OF DIMENSIONALITY

## *2.1. Typical BCI systems*

The functional structure of a typical BCI is presented in Figure 2.1 The BCI interprets the user's conscious intention by taking measurements of brain activity. These measurements, which could be electrophysiological, magnetic, or metabolic [52], are pre-processed and converted into features. Using well established methods, a subset of the most valuable feature vectors is created. These features are used for training the BCI to issue the correct control commands for operating an external system [53]. Each of the BCI components shall be explained in detail, in the following sections.

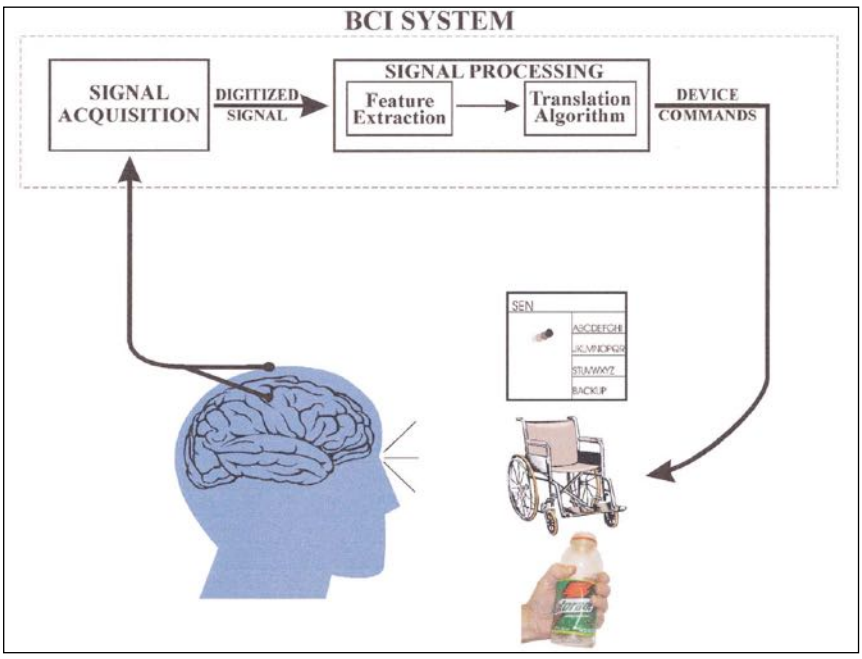



***Figure 2.1.*** *The design and operation of a Brain-computer interface [53].*

## *2.2. Data acquisition from cortical activity*

The measurements from brain activity can be made by several devices, which fall into three major categories [54]:

- Electrophysiological: such as ECoG and EEG
- Magnetic: such as MEG
- Metabolic: such as FMRI, and FNIRS

Details of each measurement system are presented in the following sections.

### *Electrocorticogram (ECoG)*

Cortical Neural activity: since the 60s, there have been many attempts to use implants, which directly connect to the human nervous system. The purpose of implanting these devices into the human brain varies from stimulation of the brain to remedy movement issues or modify the mood for patients with depression. Deep Brain Stimulation (DBS) has also been used to help people with Parkinson's disease, dystonia, or Essential tremors [55]. Implants could also be used for diagnosis and treating epilepsy [56-58]. From a BCI design point of view, a clear advantage of having direct access to the brain via an Eco implant, is avoiding the scattering effect of the skull and scalp tissue on the brain signal leading to much greater signal to noise ratio. ECoGs are not as susceptible to EMG and EOG artifacts. Also, implants could provide greater spatial resolution and reduce time delays [59].

An important invasive ECoG-based investigation by Leuthardt et al., 2011 [56], in which a one-dimensional computer cursor was controlled (left and right) with accuracies between 68% and 91%. The importance of the High-Gamma range was emphasized greatly in this work. The electrode array of 64 electrodes (8x8) 10mm apart. A series of four phonemes ('oo', 'ah', 'eh', and 'ee') or rested. Were used for binary classification. The data recorded during this study were converted to the frequency domain by autoregressive spectral estimation in 2 Hz bins ranging from 0 to 550 Hz. For each electrode and frequency bin, candidate features were chosen by calculating the coefficient of determination (r^2 ) between the 'rest' spectral power levels and the activity spectral power levels for each phoneme, and between spectral power levels for all possible phoneme combinations. Those ECoG features (Electrode and frequency bin) with the highest r^2 values, i.e. the features that had most of their variance explained by the task, were chosen as control features. 99 Trials were recorded for covert speech tasks. Classification accuracy was calculated using online performance.

The next study of note is by Pei et al., 2011a [60], which defined the most important cortical areas in word production. Each participant had a maximum of 64 electrodes implanted. 36 words was composed of one of four different vowels (/ɛ/, /æ/, /i:/ and /u:/) and one of nine consonant pairs (i.e. /b_t/, /c_n/, /h_d/, /l_d/, /m_n/, /p_p/, /r_d/, /s_t/, /t_n/). 140 trials were recorded for each subject. Then, separately for each channel and analysis (overt or covert speech, vowels or consonants), we ranked the ECoG Then, separately for each channel and analysis

(overt or covert speech, vowels or consonants), we ranked the ECoG features using the MRMR (maximum relevance and minimum redundancy) criterion [57]. We submitted the best (35 or 40 for decoding consonants or vowels, respectively) features at each location to a Naive Bayes classifier and used the optimized features to decode from each trial the vowel and consonant pair group for the target word of that trial, respectively. The average classification accuracy for 4-class covert speech was 38%.

Later, Pei et al., 2011b [48] took this work to the next stage with focusing on the Spatio-temporal dynamics of word production. In this study 9 patients were implanted with up to 64 electrodes. 36 words of the nature consonant-vowel-consonant were used to record covert speech tasks. The small cortical activations over primary motor cortex indicated that the dominant processes in covert word production were word comprehension and phonological processing rather than imagery related to the motor actions of speech production.

The next important work by Ikeda et al., 2014 [61], used ECoGs to study the role of Broca's area in pre-processing of vowels. Three vowels /a/, /i/, and /u/ were selected for classification. A total of 90 trials were recorded for each user. For each electrode, the power spectra in the high-Gamma bands were extracted from ECoG signals in the blank period of individual trials using fast Fourier transform (FFT). The power spectra were used as feature vectors for decoding analyses. Classification was performed using 15-fold cross validation and SVM classifiers. The cortical areas able to discriminate single vowels were the primary

motor area, the premotor cortex, STG, and so on. Decoding accuracies from these areas ranged from 42.2 to 46.7% for three classes.

Later the same year, Martin et al., 2014 [62] conducted another invasive study and identified the critical frequency range of 70-150 Hz and emphasized the roles of the Wernicke and Broca's areas in covert speech production. In this study 16 electrodes were implanted. The nursery rhyme "Humpty dumpty" was spoken overtly and then decoded covertly using the overt speech as baseline. The recognition rate was significantly greater than chance level using this approach.

The next significant study is by Martin et al., 2016 [63] using ECoG recordings and classifying covert word pairs. The importance of the High-Gamma frequency bands and the involvement of the temporal, frontal, and pre-motor areas in word production are discovered. For binary classification, the six words 'spoon', 'cowboy', 'battlefield', 'swimming', 'python', and 'telephone' were recorded 24 times each. Time Wrapping was used to generate features. Using SVM classifier a word-pair classification accuracy of 88% was reached in this study.

In 2018, Ibayashi et al. [64] Conducted a study using a combination of single/multi-unit activity (SUA/MUA), local field potential (LFP), and electrocorticography (ECoG) signals. Feature vectors were built from spike frequency acquired from SUAs and event-related spectral perturbation derived from ECoG and LFP signals. The results indicated that the accuracy for five monophthongal vowels (/a/, /i/, / β /, /e/, and /o/) was maximum when features

from multiple signals were merged and optimized for each participant and reached 59% when averaged for six participants. A silicone-coated electrode array consisting of a combination of six microneedles and three macroelectrodes was used. Spike firing frequency and LFP and ECoG spectral power for the different frequency bands (Alpha 8–12 Hz, Beta 14–30 Hz, Delta 2–4 Hz, Theta 4–8 Hz, low-Gamma 30–80 Hz, high-Gamma 80160 Hz, ultra-Gamma 160–240 Hz) epoched from 0 to 600 ms post-cue were analysed for each trial. 100 feature vectors were used for each of the 50 trials and results were found using cross-validation using a Sparse Logistic Regression classifier. The results of the study indicate using a combination of signals leads to the best results.

Also in 2018, Martin et al. [33] conducted a review on the progress and challenges in using implanted electrodes in speech BCIs. They suggested that improving task design, training the participants, discovering behavioural markers, speech recognition models, increasing data, implementing unsupervised learning, and improving the design of electrodes may improve covert speech BCIs.

In 2019, Rabbani et al. [65] performed a comprehensive review and study on different aspects of speech production and decoding (semantic, auditory, and articulatory), increasing the frequency range of interest for linguistic tasks from 70-170Hz to 70-350Hz within the first 350ms post onset. The advantages and problems of using long-term subdural implants are studied and the possibility of grater populations using such BCIs are discussed.

***Table 2.1:*** *The most significant ECoG studies on covert speech*

| *Publication* | *Classes* | *Nature of classes* | *Trials* | *Features* | *Classifier* | *CA%* |
|---|---|---|---|---|---|---|
| Leuthardt 2011 | 5 | oo, ah, eh, ee, rest. binary classification | 99 | R^2 | Statistical | 91% |
| Pei 2011a | 4 | ɛ, æ, i: and u: | 140 | 40 | Naive Bayes | 38% |
| Pei 2011b | 36 | Consonant-vowel-consonant | 140 | R^2 | N/A | N/A |
| Ikeda 2014 | 3 | a, i, and u | 90 | N/A | SVM | 46% |
| Martin 2014 | N/A | Nursery rhyme overt and covert | 3 | N/A | Statistical | N/A |
| Martin 2016 | 6 | Spoon, cowboys, battlefield, swimming, python, telephone | 24 | N/A | SVM | 88% |
| Ibayashi 2018 | 5 | spoken vowels (a, i, β, e, and o) | 50 | 100 | Sparse Logistic Regression | 59% |
| Martin 2018 | N/A | Review study | N/A | N/A | N/A | N/A |
| Rabbani 2019 | N/A | Review study, pushing the envelope | N/A | N/A | N/A | N/A |

## *Electroencephalogram (EEG)*

EEG is the most commonly used system in BCIs [66-75]. EEG is non-invasive, provides real-time measurements of brain activity, and is relatively cheap [53]. EEG uses several electrodes (64, 128, or 256), which are places over the scalp and electrically connected by conductive gel. The downside of EEG is limitation of spatial resolution due to the smearing effect of the tissue between the cortical surface of the brain and the electrodes. The hardware, which is used by the BCI research group at the University of Essex, is the BioSemi ActiveTwo system [76].

***Figure 2.2.*** *the 256-channel BioSemi ActiveTwo A/D box with battery [76].*

There are three major types of cortical activity that are measured using EEG. These include:

- Evoked potentials [77], such as Steady State Visual Evoked Potentials (SSVEPs) [78], Flash Visual Evoked Potentials (FVEPs) [79], and the P300 [80-84]. VEPs can be measured by EEG, over the visual cortex. BCIs have used these signals to determine the direction of gaze to control a computer mouse on a screen [78-80, 85]. P300 evoked potentials appear 300ms after a significant stimulus, among routine stimuli, and a positive peak can be detected by EEG, over the parietal cortex. These "odd-ball" stimuli may be auditory, visual, of somatosensory. These signals are difficult to detect.
- Slow cortical potentials (SCPs): low frequency voltage changes that shift over 0.5-10sec and can be recorded by EEG over the scalp. Negative SCPs are associated with movement and positive SCPs are an indication of reduced cortical activity [53, 54, 66, 86, 87].
- Mu and Beta rhythms: Brain activity is usually divided into frequency bands: in people that are awake, the sensory and cortical areas produce an 8-12Hz frequency when they are inactive. The mu rhythm focuses on the somatosensory and motor cortex, while the alpha rhythm is focused over the visual cortex. Mu rhythms are accompanied by 18-26Hz beta frequencies. When movement begins, there is a decrease in mu and beta rhythms also known as event related de-synchronization (ERD). With relaxation, there is an increase in mu and beta rhythms also known as event related synchronization (ERS).

These changes do not require actual movements and imagination of movement also produces the same effect [53, 86, 88]. Table 2.2 contains the most recent examples of EEG research on covert speech and their use in BCIs [89].

One of the earliest examples of using Covert speech in non-invasive BCIs is D'Zmura et al., 2009 [7] where imagined speech of two syllables are spoken in one of three rhythms and was discovered that the Alpha, Beta, and Theta bands contain significant information. Two syllables "ba" and "ku" in three different rhythms were used to create six conditions. 120 trials were recorded per condition. EEG was recorded using a 128-channel system. Hilbert transform was used to create envelopes. The envelopes serve as input to the matched-filter classification and as data used to construct the matched filters. An electrode's average envelopes, found by averaging across trials for each of the six conditions, serve as matched filters. This is done by finding the pseudoinverse the average envelopes for the six conditions for each electrode. The highest classification accuracy was achieved in the beta band (80%).

The next significant work is Porbadnigk et al. 2009 [90], where EEG was recorded with 16 electrodes. Five words alpha, bravo, charlie, delta, echo, are presented in blocks and classified using an LDA classifier to achieve a classification rate of 45.5% . However, when using 20 randomised trials, the performance dropped to chance level. This was due to the inconsistency of the experimental paradigm. This problem was avoided in my current work.

The next important work is Brigham and Kumar 2010 [16], where 7 volunteers imagined two syllables BA and KU. Autoregressive coefficients were used as features and a K-nearest Neighbour classifier was able to distinguish these classes. EEG was recorded using 128 channels, and a maximum of 14 trials were recorded for each class. The best accuracy achieved was 67%.

The Next noteworthy research is Deng et al. 2010 [91], where two syllables ba and ku were covertly spoken in one of three rhythms. EEG was recorded using 128 channels. 120 trials were recorded for each of the six conditions. 9 features per channel were extracted using Hilbert-Haung transform and it was shown that rhythmic structure can be successfully detected using EEG. Using LDA classifiers, a maximum accuracy of 26% was achieved for six-class classification.

The next study come in 2013 by Kim et al. [21] , where two categories of meaning (face/number) were classified using support Vector Machines achieving accuracies of 71.68%. Thirty channels, 49 frequency bands, and 80 time-steps were used as JTF features. When the number of features were reduced to 1928 from the total of 117600 using SVM based recursive feature elimination, the classification accuracy was increased to 92% for two meaning categories.

The next important study by Song and Sepulveda 2014 [92], used 64 EEG sensors to focus on detection of imagined high pitch tone production in inhibited and imagined speech for the purpose of onset detection. 40 trials were recorded for each user. Autoregressive model of order 6 was used for feature generation. The highest classification of 85% was achieved using LDA classifiers.

The next important work is Iqbal et al. 2015 [15], which classified between sounds /a/ and /u/, and idle state achieving classification accuracies between 77%-100% using SVM classifiers. EEG was recorded using 4 electrodes over the motor cortex. 50 trials were recorded for each class. Features were chosen as the variance, entropy, and signal energy in the normalised frequency range of 0.5 to 0.9 for the four channels (12 features in total).

In 2015, Zhao et al. [93] managed to classify 7 different phonemic/syllabic prompts iy, uw, piy, tiy, diy, m, n, in 3-class classification reaching up to 95% accuracy with Deep Belief Network classifiers. EEG was recorded using 62 electrodes. 12 trials were recorded per class. Various features were generated including the mean, median, standard deviation, variance, maximum, minimum, maximum ± minimum, sum, spectral entropy, energy, skewness, and kurtosis, and the first and second derivates of the above features. This results in 1197 features for each channel of the trial, for a total of 65,835 features across the 62 channels. Features were ranked by their Pearson correlations with the given classes for each task independently and we select the N features with the highest correlation coefficients, where $N \in [5..100]$. The proposed DBN classifier outperformed baseline SVM classifiers. Later the next year, Iqbal et al. 2016 [94] took this work further by including the time domain analysis of their previous study and using features such as mean and standard deviation. The maximum classification accuracy of 100% was achieved.

The next important work is by Yoshimura et al. 2016 [95] using a combination of EEG and FMRI on classification of Japanese vowels /a/ and /i/ and enhancing classical EEG with FMRI forming EEG-estimated cortical currents. EEG was recorded using 32 electrodes. 50 trials were recorded per task. Hyper-parameters were used for classification after estimating EEG cortical currents from EEG signals. Sparse Logistic Regression (SLR) classifiers achieved a maximum classification accuracy of 55% compared to 40% for EEG.

The next work is by Gonzalez et al. 2017 [96], where five covert speech tasks Left, Right, Up, Down, Select, were used in classification of EEG, sonified EEG, and textified EEG. Signals were recorded using 14 channels. DWT features were extracted. Each word was recorded 33 times. Using three classifiers, RF, SVM, NB, the three types of data were classified. The greatest accuracy was achieved with EEG represented as text with accuracy of 89%.

After this, the next important study is by Nguyen et al. 2017 [12], where three categories of short words, long words, and vowels (In, out, up, cooperate, independent, a, i, u) where a novel method based on covariance matrix descriptors, which lie in Riemannian manifold, and the relevance vector machines classifier is proposed. EEG was recorded using 64 electrodes and 100 trials per class were recorded. CSP was used to generate features and 35 features were used in classification using Relevance Vector Machines (RVM) classifiers. The highest classification accuracy achieved for binary classification is 90% and for

three classes 70%. RVM outperformed SVM due to its Bayesian leaning component.

The next work is by Rezazadeh et al. 2017 [97], where two words YES, and NO and a rest state were classified using Multilayer Perceptron (MLP) ANN classifiers. EEG was recorded with 64 electrodes and 60 trials were recorded for each class. DWT was used to extract features. The root-mean-square (RMS) and standard deviation (SD) of the outputs from each DWT decomposition level were used as features for classification. A total of 496 DWT features were generated from each trial (62 electrodes x 4 decomposition levels x 2 features, i.e., SD and RMS). For binary classification of speech vs. Rest a maximum classification of 85% was reached. Classification of yes vs. No had a maximum of 69%. For three class classification a maximum of 63% accuracy was reaches. The MLP outperformed many traditional classifiers using the data in this work.

The next important work is by Hashim et al. 2018 [98], where two word classes yes and no were distinguished using dry electrodes. EEG was recorded using 14 channels, 6 of which were used in the analysis. For each trial, Mel Frequency Cepstral Coefficients (MFCC) are extracted and used as features. Using KNN classifiers a maximum classification accuracy of 63% is reached using the dry electrode EEG system.

Cooney et al. 2018 [11], classified the impressive number of 11 covert speech classes (iy, piy, tiy, diy, uw, m, n, pat, pot, knew, gnaw). EEG was recorded from 64 electrodes. Each class was recorded 12 times. Mel Frequency

Cepstral Coefficient (MFCC) features were compared with the performance of linear and non-linear features. Each of the 13 MFCCs calculated for all 62 channels and all 17 data windows results in a total of 13,702 features which are used as input training features to the classification models. The highest classification accuracy was achieved using SVM classifier for 11 classes and achieved the impressive accuracy of twice the chance level of 19.6%.

In 2018, AlSaleh et al. [99] conducted a study to classify five covert words ("Left", "Right", "Up", "Down" and "Select"), using four different classifiers (Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), and Linear discriminant analysis (LDA)) using two experimental settings. The first where the start and end of trials are determined by user's mouse click, and the second where start and end within a set time. EEG was recorded using a 12-channel Emotive system and 144 feature vectors were extracted per trial using Discrete Wavelet Transform with 6 levels of decomposition. 100 trials were recorded in blocks, and 35 were used for training. The greatest result of 80% were obtained using RF classifier and 3.5-4 seconds of fixed trial duration.

In 2019, Sereshkeh et al. [100] conducted a study combining EEG and fNIRS. Three tasks of covertly speaking "yes", "no", and rest were classified. 132 trials were recorded using 30 EEG channels and 40 fNIRS channels. 180 DWT features were used. Classification was with LDA and estimation of performance was achieved with 10-fold CV. The performance of the system using only EEG

compared to using both signal types was improved from 69% to 76%. The results demonstrate improvement in BCI performance when both modalities are used.

In 2019, Cooney et al. [101] performed a novel experiment using Convolutional Neural Networks to classify word pairs. The performance of deep CNN, and shallow CNN has been compared with LDA classifier trained with filter-bank common spatial patterns features FBCSP focusing on band power differences between classes. Nested cross-validation was used to permit optimization of two hyper-parameters. For FBCSP the first of these is the number of selected spatial filter pairs (1,3,4,5). The second hyper-parameter used here was the mutual information quantization level, with the values considered being 6, 8, 10 and 12. For the CNNs the hyper features were learning rate and the number of filters implemented in the final convolutional layer. participants imagined the production of six Spanish words: “arriba”, “abajo”, “derecha”, “izquierda”, “adelante” and “atrás”. EEG was recorded using 6 channels and a frequency range of 2-40 Hz was used. 50 trials per class were recorded. The results show that deep CNN outperforms rLDA from 57% to 65%. This approach shows great promise for future work.

Also in 2019, Cooney et al. [102] used CNN and two transfer learning methods to improve generalizability of covert speech tasks in EEG data. Both TL approaches involved conditional training of the CNN on all participants, excluding the target subject. A subset of the target participant data was then used to fine-tune either the input or output layers of the CNN. Participants imagined

speaking the five vowels “/a/”, “/e/”, “/i/”, “/o/” and “/u/”. EEG signals were recorded using an 18-channel Grass® Analog amplifier. The original dataset was filtered between 2 Hz and 40 Hz. The feature extraction section of the network consists of six convolution layers. The input of the CNN consists of two convolutional layers, the first to perform temporal convolution and the second for spatial filtering. This construction has been conceived of as a feature extraction stage analogous to that of FBCSP and is designed to decode band power features from EEG. The feature map obtained from the feature extraction stage is passed to the final block for classification. 5-fold cross-validation scheme was applied to split the data into training, validation and test sets. Two different TL methods were tested. TL1 involved training the CNN on all source subject data before target-subject data was used to finetune the input layers of the CNN. TL2 employed the same training strategy, but fine-tuning was implemented on later layers of the CNN. These TL methods were compared to a non-TL approach to training with the same CNN architecture. Both TL approaches outperformed the baseline classifier and suggest the possibility to improve the generalization of covert speech tasks for BCIs.

In 2019, Saha et al. [103] conducted a study on classification of 7 covert phonemes /iy/, /piy/, /tiy/, /diy/, /uw/, /m/, /n/, using Hierarchical Deep Learning. The proposed network is composed of hierarchical combination of spatial CNN and temporal CNN (TCNN) cascaded with a deep autoencoder. For the phonological categorization task, input data for CNN and TCNN (covariance

matrix) is of length 61×61 and 1,891 respectively, while for the speech recognition task, the input data (phonological features) is of length 6 × 256 and 1,536 respectively. The input data for deep autoencoders pertaining the two tasks is of length 2,915 (1,891 TCNN + 1,024 CNN features). EEG signals from the KARA database were used. For binary classification, a maximum classification accuracy of 83% was achieved.

In 2020, Imani et al. [104] used ICA for classification of imagined speech. EEG signals were recorded using a 19-channel Micromed helmet in unipolar mode. The four directions up, down, left, and right were recorded six times per class. Classification was achieved using ANN classifiers with 20 neurons and two layers. 10-fold cross validation was used and a maximum accuracy of 60% was achieved.

In 2020, Bakhshai et al. [105] conducted a study on EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density. CSD matrices are evaluated for EEG signals obtained from different channels, and the distances between these matrices are considered as measures for imagined speech recognition. In this work, channel selection and frequency band detection during imagined speech is evaluated with statistical methods. Riemannian geometry is used as the framework of feature extraction. In this research study, a novel CSD-based Riemannian distance is presented, and it is used as a feature for classification of EEG signals during imagined speech. Non-parametric cluster-based permutation test is employed to determine the

statistical significance of the extracted features and control the false positive rate caused by multiple comparisons. The permutation test uses a test statistic that is based on clustering of adjacent spatial locations that exhibit a similar P feature (sum of CSD values in each frequency band) difference between imagined speech and baseline. CSD matrices were considered as inputs to the algorithm and their Riemannian distances were considered as features. EEG was recorded from 64 channels. Imagined speech of either of the four words /gnaw/, /knew/, /pot/, /pat/ were used as classes. A total of 43 trials per class were recorded. Using KNN classifiers in binary classification, a maximum classification accuracy of was achieved.

The next significant study is by Krishna et al, 2020 [106] where 9 sentences used for identification from continuous EEG recordings using a long short-term memory (LSTM) based regression model and Generative Adversarial Network (GAN) based model. Wasserstein generative adversarial networks (WGAN) to decode the Mel-frequency cepstral coefficients (MFCC) features of the audio that the subjects were listening from the EEG signals which were recorded in parallel while they were listening to the audio as well as we decode MFCC features of the sound that the subjects spoke out from the EEG signals which were recorded in parallel with their speech. The sentences were covertly spoken 3 times for each user and 150 features were extracted. A maximum of 10% for word recognition was achieved in this study using continuous EEG recordings.

***Table 2.2:*** *Important EEG studies on covert speech and its use for BCIs.*

| *Publication* | *Classes* | *Nature of classes* | *Trials* | *Features* | *Classifier* | *CA%* |
|---|---|---|---|---|---|---|
| D'Zmura 2009 | 6 | ba, ku, three rhythms, binary classification | 120 | 768 | Matched filter | 87% |
| Porbadnigk 2009 | 5 | alpha, bravo, charlie, delta, echo | 20 | N/A | LDA | 19% |
| Brigham 2010 | 2 | ba, ku, binary classification | 14 | 128 | KNN | 67% |
| Deng 2010 | 6 | ba, ku, three rhythms, 6-class classification | 120 | 1152 | LDA | 26% |
| Kim 2013 | 2 | Numbers category, face part category | N/A | 1928 | SVM | 92% |
| Song 2014 | 2 | High-pitch tone vs. rest | 40 | 64 | LDA | 85% |
| Iqbal 2015 | 3 | a, u, rest | 50 | 12 | SVM | 100% |
| Zhao 2015 | 7 | iy, uw, piy, tiy, diy, m, n, 3-class classification | 12 | 100 | DBN | 95% |
| Iqbal 2016 | 7 | iy, uw, piy, tiy, diy, m, n, 3-class classification | 50 | 8 | statistical | 100% |
| Yoshimura 2016 | 3 | a, i, rest | 50 | Hyper | SLR | 65% |
| Gonzalez 2017 | 5 | Left, Right, Up, Down, Select | 33 | N/A | RF, SVM,NB | 89% |
| Nguyen 2017 | 8 | In, out, up, cooperate, independent, a, i, u | 100 | 35 | RVM | 70% |
| Rezazadeh 2017 | 3 | Yes, no, rest | 60 | 496 | MLP | 63% |
| Hashim 2018 | 2 | Yes, no | 50 | N/A | KNN | 63% |
| Cooney 2018 | 11 | iy, piy, tiy, diy, uw, m, n, pat, pot, knew, gnaw | 12 | 546 | SVM, DT | 20% |
| AlSaleh 2018 | 5 | Left, Right, Up, Down, Select | 35 | 144 | Random Forest | 80% |
| Sereshkeh 2019 | 3 | yes, no, and rest | 132 | 180 | LDA | 77% |
| Cooney 2019 | 6 | arriba, abajo, derecha, izquierda, adelante ,atrás | 51 | Hyper | rLDA, CNN | 66% |
| Cooney 2019 | 5 | a, e, i, o and u | 100 | Hyper | TL, CNN | 39% |
| Saha 2019 | 7 | iy, piy, tiy, diy, uw, m, n | 11 | 2915 | TNN, TCNN | 83% |
| Imani 2020 | 4 | Up, down, left, right | 6 | 20 | ANN | 60% |
| Bakhshai 2020 | 4 | gnaw, knew, pot, pat | 43 | N/A | KNN | 90% |
| Krishna 2020 | 23 | 9 sentences, continuous EEG | 3 | 150 | Statistical | 10% |

### *Magnetoencephalogram (MEG)*

MEGs measure the magnetic activity of the brain and provide high temporal resolution like EEGs [107-110]. MEGs use superconductors and are bulky, expensive, and difficult to use [111]. The spatial resolution of MEGs is limited due to the smearing effects of the tissue between sensors and the cortical surface of the brain.

***Functional Magnetic Resonance Imaging (FMRI)***

FMRIs measure changes in the hemodynamic response of the brain with excellent spatial resolution. However, the temporal resolution of this system is poor. FMRIs are bulky and expensive and require expert technicians [111-114].

***Functional Near-Infrared Spectroscopy (FNIRS)***

FNIRs provide a spatial map of the brain's functional activity based on hemodynamic changes. This system is non-invasive and portable and has been successfully used for motor imagery BCIs [52, 115-118].

## *2.3. Pre-processing*

The acquired data is pre-processed before extracting features. The purpose of this stage is to maximize signal to noise ratio, and ensure the data has the correct temporal and frequential content. For EEG data, pre-processing may include down-sampling, referencing, filtering, artifact removal, and epoch selection. A notch filter can remove the power line noise (50Hz) and a band pass filter can select the appropriate frequency range for analysis. The most common referencing methods are common referencing (additional channels at mastoids), bipolar referencing, common average referencing, and the surface Laplacian [119]. The main sources of noise in EEG recordings are muscle activity (EMG) and eye blinks (EOG). These artefacts can be removed using ICA, CSP, PCA, or

may be rejected by visual inspection of recorded data [87, 119-121]. The entire pre-processing stage can be performed with the use of the open source MATLAB™ toolbox EEGLAB [122], which includes automatic artifact removal utilities [123].

## *2.4. Feature generation*

The purpose of feature generation is to extract detailed spatial, temporal, and frequential information from the data to identify the user's intention [53]. Many methods have been used for feature extraction from EEG signals [119, 124-126]. In addition to artefact removal, PCA [127], ICA [128], and CSP [13] have been successfully used for feature extraction. Among these three methods, only Common Spatial Patterns [13, 17-19, 22] maintains electrode location information. PCA and ICA lose the connection to EEG channel location. Wavelet Transformation (WT) is commonly used for feature extraction [129]. This joint time-frequency method used a variable and adaptable analysis window, to maintain time information for high frequencies, and accurate frequency information for low frequencies. Band Power features can be generated using a Fourier Transform, and contain information about signal energy levels in different frequency bands [130].

Time information can be included to such features by performing the Discrete Fourier Transform (DFT) for overlapping time windows. This method

is known as Power Spectral Density (PSD) [131]. The Gabor transform [132] is a windowed Fourier Transform using a Gaussian sliding window. This method preserves the information about channel locations in the data and provides frequency band information and time step information. Autoregressive coefficients [133] require a shorter time period to record EEG and provide good frequency resolution [134].

## *2.5. Clustering in high dimensional spaces*

The purpose of feature selection is to identify a subset of features that achieve the best classification accuracy with high computational efficiency [135]. There are three general methods for feature selection: the Filter approach, the Wrapper approach, and hybrid methods [136]. Filter methods evaluate the goodness of features independent of the type of used classifier. For example, the Davies-Bouldin index ranks the features based on a distance measure [137]. Wrapper methods use a pre-defined classifier to test the classification accuracy of a selected subset. This approach has a significantly higher computational cost compared to the filter approach [138].

A solution to this problem is using a combination of filters and wrappers. This approach is known as the hybrid approach [139]. For example, a fixed size subset of best features identified by a filter, can be further optimized using sequential selection [133].

This raises a problem known as “the curse of dimensionality,” that was first identified by Richard Bellman, in his book on control theory [140]. High dimensions create problems for data mining methods, and to clustering specifically. The most important problem in high dimensions is that as the data becomes very sparse, there is a loss of discriminative ability [141] for distance, proximity, or density measures that would work well for low dimensional spaces [142].

Another problem arises when irrelevant attributes exist among the features which is known as the correlation clustering problem. This problem can be avoided with the use of subspace clustering methods [143]. Correlation Clustering is associated with feature vector of correlations among attributes in a high dimensional space. These are consistent to guide the clustering algorithm. Depending on the nature of the data, one clustering algorithms may be more suitable than others and no algorithms is always superior to the others.

## *2.6. Clustering quality and clustering based feature selection*

The quality of a clustering result can be measured with a Clustering Validation Index (CVI). The purpose of a CVI is to estimate the most suitable number of clusters based on the compactness and separation of the clusters [144]. Validation indices can be divided into the three categories of internal, external, and relative. CVIs can also be used for feature selection by identifying the most

class separable feature vectors based on their indexes. The validation indices measure high similarity within clusters (intra) and high separability between clusters (inter), for which lower and higher values are preferred, respectively.

Examples of such a filtering-based approach are The correlation-based filter selection (CFS) method, The fast correlation-based filter (FCBF), and the minimum redundancy maximum relevancy (mrMR) method [145].

Among the most recent algorithms Affinity Propagation (AP) and Maximal Information Coefficient (MIC) are noteworthy [146]. Affinity propagation forms clusters using messages exchanged between data points. Given the similarities of each two distinct data points as input, AP algorithm considers all the instance as potential centroids at the start, then combines small cluster into larger ones, step by step. Maximal information coefficient (MIC). With innovative idea, they show that MIC could capture a wide range of associations both functional and not. Furthermore, the value of MIC is roughly equal to the coefficient of determination R2 in statistics. Both MIC and AP require labelled information.

A combination of these methods called MICAP can be used for unsupervised learning with no labels and directly finds the key attributes from the data. MICAP makes features with high dependence cluster together keeping only the centre feature of each cluster. The algorithm follows a simple idea that takes the MICs as the relationship metric for each pair of features, and clusters them using the affinity propagation algorithm.

In this work, we classify four covert speech tasks. In the first experimental setting presented in chapter three, EEG data is recorded with 20 electrodes. Ten trials were recorded for each class. Using the Gabor transform, EEG data from all channels for one trial is converted into 81,920 **Gabor features** (20 channels x 64 frequency-bands x 64 time-steps) as seen in figure 2.3. Instances of Gabor features in different trials (value of Gabor coefficient in that trial) are referred to as **vectors**.

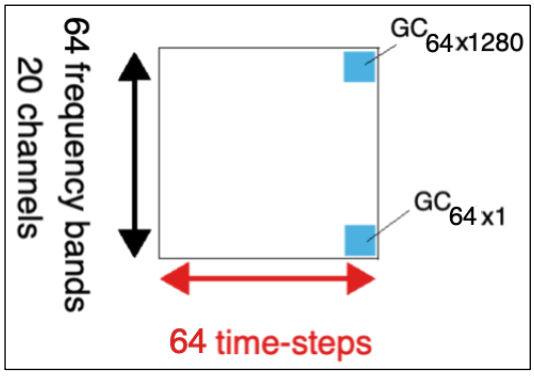



***Figure 2.3:*** *The Gabor feature space for one trial as implemented in chapter three. EEG data from 20 channels for one trial is converted into 81,920 Gabor features (20 channels * 64 frequency-bands * 64 time-steps)*

Considering that the exact number of classes and their labels are known in this work, a filtering feature selection approach based on a non-parametric cluster-based permutation is used, in which Gabor features are given a Davis Bouldin index, one at a time [93, 105, 147-150]. The DBI is a function of the sum of within-cluster scatter to between-cluster separation and the most valuable Gabor features have the smallest DBI. Gabor features are sorted based on their DBI, and a subset of the most valuable ones are selected for classification.

In chapter three, ten trials are recorded per class and 8 trials are used for training. For each class, there will be 8 vectors of the particular Gabor feature, for which a mean and standard deviation are calculated. In binary classification (i.e. BA vs. FO) the DBI is found as (std(BA)+std(FO)) / (mean(BA)-mean(FO)). Figure 2.4 illustrates this approach.

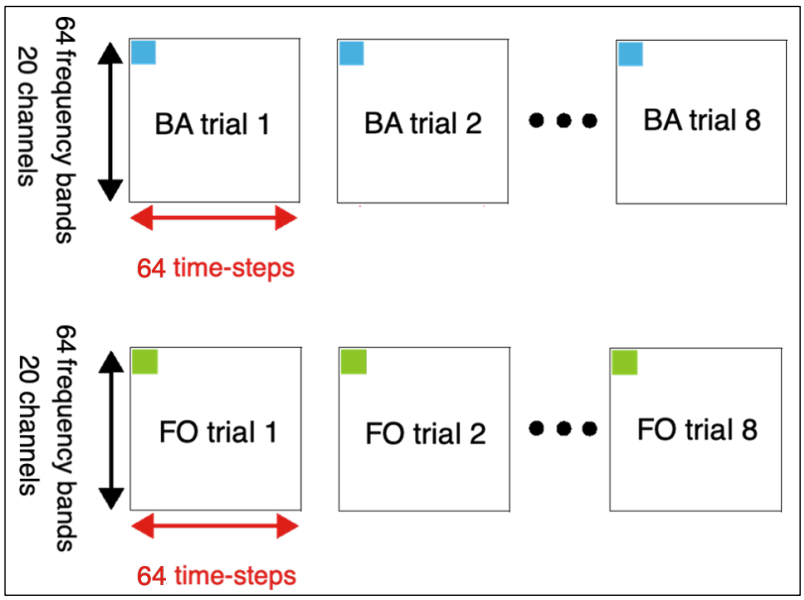
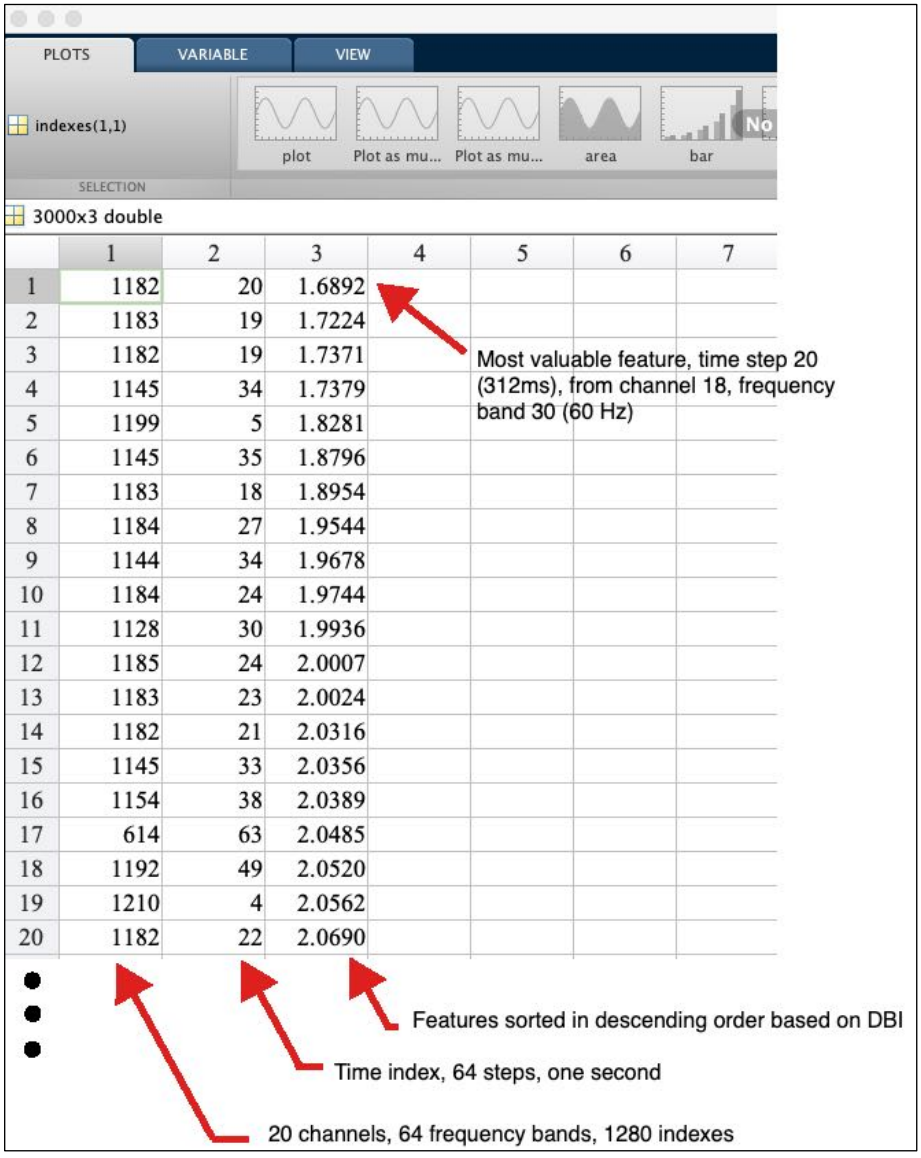



***Figure 2.4:*** *Eight out of ten trials are used for training. For binary classification, Gabor features are assigned a DBI, one at a time. In each class, the mean and standard deviation of the 8 vectors of the particular Gabor feature are calculated. The DBI is found as (std(BA)+std(FO)) / (mean(BA)-mean(FO)).*

In binary classification (i.e. BA vs. FO), for assigning a DBI to one Gabor feature there are 8 vectors (instances) of that Gabor feature (instances greater than number of features). Considering this process is conducted independently for each Gabor feature, the curse of dimensionality does not apply here, despite the very large number of total Gabor features. The most valuable features have the smallest DBI. Based on DBI, Gabor features are sorted in order of their relevance,

the indexes of the most valuable Gabor features are saved, and vectors of these Gabor features are used for training and testing. Figure 2.5 illustrates such an example. In 4-class classification, a conservative estimation of DBI is selected based on binary DBIs. This shall be explained in detail in chapter three.




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 1182 | 20 | 1.6892 | | | | |
| 2 | 1183 | 19 | 1.7224 | | | | |
| 3 | 1182 | 19 | 1.7371 | | | | |
| 4 | 1145 | 34 | 1.7379 | | | | |
| 5 | 1199 | 5 | 1.8281 | | | | |
| 6 | 1145 | 35 | 1.8796 | | | | |
| 7 | 1183 | 18 | 1.8954 | | | | |
| 8 | 1184 | 27 | 1.9544 | | | | |
| 9 | 1144 | 34 | 1.9678 | | | | |
| 10 | 1184 | 24 | 1.9744 | | | | |
| 11 | 1128 | 30 | 1.9936 | | | | |
| 12 | 1185 | 24 | 2.0007 | | | | |
| 13 | 1183 | 23 | 2.0024 | | | | |
| 14 | 1182 | 21 | 2.0316 | | | | |
| 15 | 1145 | 33 | 2.0356 | | | | |
| 16 | 1154 | 38 | 2.0389 | | | | |
| 17 | 614 | 63 | 2.0485 | | | | |
| 18 | 1192 | 49 | 2.0520 | | | | |
| 19 | 1210 | 4 | 2.0562 | | | | |
| 20 | 1182 | 22 | 2.0690 | | | | |

***Figure 2.5:*** *Gabor features are sorted based on their DBI (third column). The indexes of the most valuable subset of features (first and second column) are saved, and vectors of these Gabor features are used for training and testing. In the above example, the most valuable Gabor feature is from channel 18, frequency band 60Hand time step 625ms.*

## *2.7. Discriminant analysis in high dimensional data*

One of the most common methods in designing a classifier, is the use of supervised machine learning. In this method, a classification object is created using labelled examples of each class. The performance of this classifier is tested using new data. Many classification algorithms have been used in BCIs [117]. The performance of each algorithm depends on the type of data.

MATLAB™ provides a "Classification Learner" toolbox, which compares the performance of different algorithms. This toolbox can combine several weak classifiers to create an ensemble classification model with greater performance than any single algorithm [151]. One of the most common classifiers used for high-dimensional data is Linear Discriminant Analysis (LDA). Linear discriminant analysis is asymptotically optimal and has Bayes risk when the dimension of the feature space is fixed and the number of trials are sufficiently high [152]. However, when the number of trials is much smaller than the dimension of the feature space, LDA cannot be used as LDA computes class means and the covariance matrix, which might be unknown [153], or the covariance estimates may not have complete ranking, and cannot be inverted [154]. One solution to this problem is the use of Genetic Algorithms, which takes in an initial population, evaluates the fitness, selects fittest values, performs mutation, and finally makes a crossover to produce the next generation [155].

Another approach to the problem of small number of trials is to use a naive Bayes version of the LDA, which is related to the diagonal LDA and assumes features to be uncorrelated [156] .

However, the most common approach for using LDA in high-dimensional data is the Pseudo LDA [157, 220, 221, 222] and also used in this work for supervised learning and classification. If the training set has C classes, and $x_{ij}$ is a d-dimensional matrix (j-th sample from i-th class) then the within class and between-class scatter matrices are presented as:

$$S_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(x_j^i - m_i)(x_j^i - m_i)^T = H_w H_w^T \quad (2.1)$$

$$S_b = \frac{1}{N}\sum_{i=1}^{C} N_i(m_i - \bar{m})(m_i - \bar{m})^T = H_b H_b^T \quad (2.2)$$

where $m_i$ , $H_w$ , and $H_b$ are defined as:

$$m_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j^i \quad (2.3)$$

$$H_w = \frac{1}{\sqrt{N}}[x_1^1 - m_1, \ldots, x_{N_1}^1 - m_1, \ldots, x_{N_C}^C - m_C] \quad (2.4)$$

$$H_b = \frac{1}{\sqrt{N}}[\sqrt{N_1}(m_1 - \bar{m}), \ldots, \sqrt{N_C}(m_C - \bar{m})] \quad (2.5)$$

$\overline{m}$ is the mean sample of the entire training set, $N_i$ is the number of training samples for the i-th class and $N$ $(=N_1+ N_2+ \ldots +N_c)$ is the total number of training samples from these $C$ classes. To calculate pseudo inverse $S_w^+$, the PLDA finds the Singular Value Decomposition of $S_w$ as:

$$\boxed{S_w = Q_1 \Lambda Q_1^T} \quad 2.6$$

where $\Lambda=diag(\lambda_1, \lambda_2, \ldots, \lambda_k)$ contains the positive eigenvalues of $S_w$, $k$ (usually equals $N$-$C$) is rank of $S_w$ and $Q_1$ consists of the eigenvectors of $S_w$ corresponding to the $k$ positive eigenvalues. Then, the pseudo inverse of $S_w$ is:

$$\boxed{S_w^+ = Q_1 \Lambda^{-1} Q_1^T} \quad 2.7$$

Then the PLDA algorithm finds the eigenvectors of $S_w^+ S_b$ corresponding to the positive eigenvalues as the projection vectors. In this work, PLDA is used for classification. In order to determine if and how the curse of dimensionality is affecting classification, different sizes of Gabor feature subsets (20, 50, 100, 300, 1000, ..., 4000) were used to demonstrate how performance is affected as a result.

## *2.8. Limitations of cross-validation*

With offline datasets, a common method for testing the performance of an analysis pipeline and a classifier is to use cross validation. The most common scheme is 10-fold cross validation, in which the data is divided into ten equal portions, and each time one portion is set aside only for testing and the training is performed using the rest of the data. The average performance of these ten folds is used to estimate classification accuracy. Using CV to estimate an error rate for a classifier which has also been trained using CV produces highly biased estimate of the true error. The correct approach for using CV for estimating true error of a classifier trained by a consistent analysis pipeline is for all steps of the pipeline, including classifier parameter tuning, be repeated in each CV loop [158].

The bias in the estimated error found by cross validation is generally reduced if a larger sample size is used for training. studies based on predictive modelling require larger sample sizes compared to standard statistical approaches. As the sample size of the simulated data goes up, the estimation bias is significantly reduced [159].

In chapter four, where block design is used, in addition to 10-fold cross validation, HV cross validation is used which separates test trials from training trials with number of buffering trials. The exact details are presented in chapter four.

## *2.9. Summary*

Based on the literature presented in this chapter, the following points shall be addressed in the following chapters:

- Selecting suitable covert speech tasks and designing a novel experimental protocol, which uses state of the art discoveries on the stages of word production and the spatial, temporal, and spectral signatures of each stage. The objective is to accurately target specific linguistic activities that are consistent, yielding high classification accuracy.
- Using joint time-frequency analysis, which also maintains information on EEG channel location. To this end, the Discrete Gabor Transform is a suitable method to generate high-resolution features.
- Due to the high resolution of the Gabor features, the dimensionality of the feature space is extremely high. To avoid the curse of dimensionality, a non-parametric cluster-based permutation with the use of DBI is selected. The DBI evaluates Gabor features one at a time, independent from other features and their total number.
- Pseudo Linear Discriminant Analysis is used for classification as it has been shown to deal adequately with high dimensional feature spaces. In addition, different sizes of feature subsets will be used to determine the extent of the curse of dimensionality.

# CHAPTER THREE:

## Motor Imagery vs. Covert speech

## *3.1. Introduction*

Motor imagery (MI) is a well-established paradigm in BCIs. The low-frequency oscillations (< 35 Hz) elicited by "MI" activity, have been detectable in EEG for many decades. MI does not occur independently and is the end-result of many cognitive functions. For example, anticipating an onset cue and initiating "imagined" movement after cue recognition requires stimulus-driven executive control, with high-Gamma activity in regions such as the pre-frontal cortex [160, 161]. To take advantage of such class dependent cognitive activity [162, 163], the entire bandwidth of the EEG signal must be utilized [164] (and not only Alpha and Beta bands). Covert word production begins with high-Gamma (>70 Hz) linguistic processing stages [40, 41, 48], followed by motor imagery of articulation [27, 34]. Language is exceedingly more complex than movement [165] and requires analysis with much higher resolution than traditional MI band power [166]. However, covert speech is more intuitive and natural for BCI communication compared to MI. In this chapter, four covert speech classes are selected based on meaning and phonetic properties. Four MI tasks in hands and feet are chosen. An experimental protocol and analysis pipeline is designed and evaluated by classifying both categories of mental activity: covert speech and MI. The performance of the system for each paradigm is calculated and the results are discussed. MI provides a well-established baseline to compare with covert speech.

## *3.2. Selecting covert speech tasks*

Brain-Computer Interface systems (BCIs) measure a user's brain activity by employing a range of devices and methods, to determine the intent of the user. The measured brain activity may be consciously generated by the user, or may be an unconscious neural response to stimuli. Among the different measurement methods, EEG has many advantages, namely it is non-invasive, relatively cheap, and it provides relatively high time resolution [53]. The most commonly used cognitive task in EEG based BCIs is Motor Imagery, in which the cortical somatotopic representation of different parts of the body is measured. In theses systems, the user imagines moving a limb in a specific way, to generate a command [147, 167]. Although motor imagery provides a useful way to consciously generate distinguishable brain activity, it requires user training and changes in the imagined movement usually occur in time, which could lead to frequent errors in classification [168]. In addition, some disabled users may experience difficulty with motor tasks even if they are imagery-based [169]. Other cognitive tasks [170], such as mental navigation, covert tone production, solving a multiplication problem, imagining a 3D object, and covert syllable production (speech which is internally generated, but not articulated) have also been shown to generate distinct, task specific EEG patterns for BCI use [110, 169, 171, 172]. Some reseachers have used a combination of covert syllable production and motor imagery [14, 17, 173]. In recent years, more attention has

been dedicated to BCIs based on linguistic tasks alone [7, 13, 15, 16, 18-20, 22, 92, 164, 174-180]. Using EEG recordings and a range of processing methods [181], only pair-wise classification has been successfull, and for three classes, or more, the performance is not much better than chance [179, 182, 183]. Only ECoG based BCIs have successfully classified four covert speech clases [56, 184, 185]. Considering the fact that speech is the most natural and intutitive form of human communication, and that language and cognition are closely related processes, a BCI system designed to understand commands, covertly spoken in the user's mind, is highly desirable. In addition to ease of use and intuitiveness, the manner of articulation in each individual (their accent) is consolidated over time. Similar to Motor Imagery, covert speech tasks activate the same language motor centres of their overt form [46, 47]. However, their activity is attenuated during covert speech. As a result, covert speech is produced with the same consistency as overt speech, however detecting the attenuated activity of the language motor centres might be difficult.

Figure 3.1, illustrates the functional division of the primary motor cortex, also known as the "Homunculus". Speech production is the most complex motor skill, which takes many years to learn and master. Almost half of the Primary Motor Cortex is allocated to muscles producing speech, which reflects this complexity [186]. Each word uses a unique combination of language producing muscles when articulated. The selected covert speech tasks must be chosen to

maximise these differences in the Motor Imagery stage of speech, occuring approximately 600 ms post onset [27, 34].

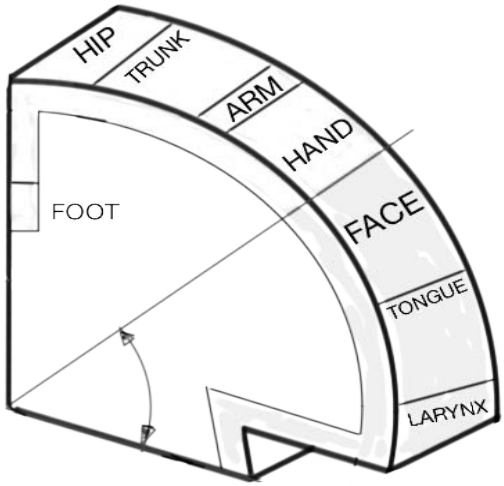



***Figure 3.1.*** *The functional division of primary motor cortex based on [186], also known as the "Homunculus". A significantly large proportion, controls muscles in the face, jaw, lips, tongue, and larynx responsible for speech.*

Dissimilar covert speech tasks also create distinctive neural activity associated with the Phonological Code Retrieval stage (Syllabification) of linguistic processing, which occurs approximately 200 ms post onset [187]. Most neocrtical territories in both hemisphers, as well as many subcortical brain regions are involved in language [165, 188]. Based on the unique cognitive neuroanatomy of each individual, the spatial, temporal, and spectral signatures may vary from person to person [39]. A linguistic BCI with four degrees of freedom, is sufficient for controling a smart device, or using a computer. In this study, the four directions (back, forward, left, and right) are shortened into "BA", "FO", "LE", and "RY" and used as covert speech tasks. These phonemic structures are used as "non-words" in the first half of the experiment. Participants are informed of the meaning

in the second half, converting the phonemic structures into “words”. Other than this, the first, and second half of the experiment are identical in every other way; making it possible to study the effects of assigning meaning on classification accuracy. These covert speech classes are cognitively appropriate directional commands, have little or no overlap with typical mind-wandering states, and provide an intuitive method of communication. For example, the user can move a cursor to the left by covertly speaking “LE”. In addition, these word classes are phonetically dissimilar, thus the motor planning for articulation is completely different for each. In order to demonstrate these differences in a quantitative manner, the properties of each consonant and vowel, such as place of articulation and manner of articulation [189, 190] are presented in figure 3.2. For example, the consonant /b/ is voiced, plosive, and bilabial.

| Consonants | /b/ | /f/ | /l/ | /r/ | | Vowels | /a/ | /o/ | /e/ | /ai/ |
|---|---|---|---|---|---|---|---|---|---|---|
| Voiced | 1 | 0 | 1 | 1 | | Lips Spread | 1 | 0 | 1 | 1 |
| Voiceless | 0 | 1 | 0 | 0 | | Lips Round | 0 | 1 | 0 | 0 |
| Plosive | 1 | 0 | 0 | 0 | | Tongue Front | 1 | 0 | 1 | 1 |
| Fricitive | 0 | 1 | 1 | 0 | | Tongue Back | 0 | 1 | 0 | 0 |
| Continuant | 0 | 0 | 0 | 1 | | Jaw Open | 1 | 0 | 0 | 1 |
| Bilabial | 1 | 0 | 0 | 0 | | Jaw Half closed | 0 | 1 | 1 | 0 |
| Labiodental | 0 | 1 | 0 | 0 | | Jaw Closed | 0 | 0 | 0 | 1 |
| Alveolar | 0 | 0 | 0 | 0 | | Diphthongue | 0 | 0 | 0 | 1 |
| Paletal | 0 | 0 | 0 | 1 | | Monophthongue | 1 | 1 | 1 | 0 |

***Figure 3.2.*** *Properties of the consonants and vowels in the word classes such as place of articulation and manner of articulation.*

## *3.3. Stages of word production*

Speech is the most natural and intuitive form of human communication. Language and cognition are closely related processes. A BCI system designed to understand commands covertly spoken in the user's mind, is highly desirable. Most neocortical territories in both hemispheres, as well as many subcortical brain regions are involved in language [165]. EEG signals can successfully identify 200-600 Hz cortical spikes [191-193] for medical diagnostic applications. In artefact-free conditions, EEG signals accurately measure induced/evoked high-Gamma brain activity, up to 150 Hz [35, 36, 38, 194]. Based on the unique cognitive Neuroanatomy of each individual, the spatial, temporal, and spectral patterns of activity may vary from person to person [39].

Word production begins with semantic (conceptual preparation), lexical (Lemma retrieval), and phonetic (phonological code retrieval and syllabification) linguistic processes, followed by planning the movements of language muscles (phonetic encoding) for articulation [27, 34].

Linguistic phonetic processing is an automatic brain function, which elicits high-Gamma (70-160 Hz) oscillations [40, 41]. In each individual, Phonetic processing activity for a specific word does not change over time [42, 43] and is not affected by priming, cognitive activity, or task frequency [44, 45]. In contrast, semantic and lexical processing, is affected by task frequency, priming, and cognitive activity [9, 10, 25], which would also arbitrarily shift the temporal

course of all following functions. These problems can be avoided by using a suitable experimental protocol.

In covert speech, the manner of articulation in an individual (their ‘accent’) is consolidated over time. Covert articulation tasks activate the same language motor centres as their overt form [46, 47]. As a result, covert speech is produced with the same consistency as overt speech. However, in covert speech, the activity of the Primary Motor Cortex is greatly attenuated [48] and may be difficult to detect by EEG. Speech production is the most complex motor skill, which takes many years to learn and master. Almost one third of the Primary Motor Cortex is allocated to muscles producing speech, which reflects this complexity [186].

Phonetically dissimilar covert speech tasks create distinctive neural activity associated with the phonological code retrieval and syllabification stages of linguistic processing [187] and involve different language muscle combinations during covert articulation. A linguistic BCI with four classes is sufficiently capable of controlling a smart device with a suitable user interface. In this study, the four directions (back, forward, left, and right) are shortened into Phonemic structures “BA”, “FO”, “LE”, and “RY” and used as covert speech tasks. These covert speech classes are cognitively appropriate directional commands, have little or no overlap with typical mind-wandering states, and provide an intuitive method of communication. For example, the user can move a cursor to the left by covertly speaking “LE”. In addition, these Phonemic structures are phonetically dissimilar.

If the word class is known by the user before the trials, the conceptual preparation stage will be completed in advance. The Lemma selection stage, with multiple competing lemmas will have temporal inconsistencies. If trials are recorded in blocks, only one Lemma is activated and selected. In block recordings, the same auditory time cue, in the form of a "beep' sound, can be used for task onset in all word classes, thus eliminating class-dependent auditory evoked responses from trials. By consolidated the semantic and lexical activities, conceptual preparation and lemma selection are complete before task onset. As a result, trials only contain automatic phonetic linguistic processing stages, and will not be affected by the temporal inconsistency of cognitive activity. Mental effort causes activation of scalp and neck muscles [195], which can mask high-Gamma cortical components. In this work, no mental effort is required from the user during trials. These conditions can be easily reproduced for the online application of this Linguistic BCI, with the same block recordings used for training.

After cue recognition (~100ms post-onset), the following stages are [48]: Lemma activation (~100-175ms post-onset), phonological code retrieval (~175-250ms post-onset) and syllabification (~250-300ms post-onset). Covert articulation (~500-800ms post-onset) and the corresponding Motor imagery activity, are separated from the linguistic stages by a ~200ms interval, during which covert articulation is designed by an internal perceptual process using the working memory and the somatosensory association cortex [39].

## *3.4. Designing experimental protocol*

In this chapter, each recording run contains four classes, which are shown in the user interface by four arrows: up, down, left, and right. Within a recording run, 10 examples of each task are presented in a random order (each run has 40 trials) to avoid user fatigue. During recording, a new task is determined by an arrow appearing on the screen for 3 seconds. After the arrow disappears, there is a 3 second standby state. Task onset is presented as a beep sound for all classes. A second beep indicates a rest period before the next trial. The experimental protocol is presented in figure 3.3.

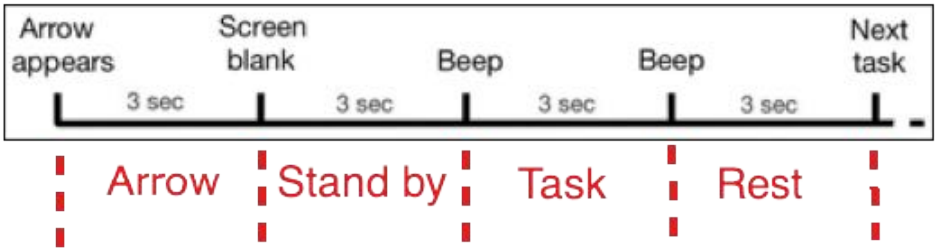



***Figure 3.3.*** The experiment protocol for recording four randomly presented trials. Each class corresponds to a directional arrow. After task presentation, a beep sound is used for all classes as task onset. A second beep indicates a rest period before the next task.

Each user completes two recording runs, which are identical in every way except for type of mental task (MI, covert speech). For MI tasks, the four arrows represent left hand movement (left arrow), right hand (right arrow), left foot (down arrow), and right foot (up arrow). In covert speech tasks, the user imagines speaking the phonemic structures: BA (back/down arrow), FO (forward/up

arrow), LE (left arrow), and RY (right arrow), which are phonetically very dissimilar tasks [196]. 1-second epochs sufficiently capture both linguistic activities and motor imagery activities for both paradigms.

## *3.5. Data acquisition and pre-processing*

Four neurologically healthy volunteers participated in this experiment. The EEG signals were recorded using an Enobio dry electrode system with 20 channels and 10/10 configuration [197]. Data was recorded at a sampling rate of 500 Hz and saved in “gdf” format. Compared to wet electrode systems, setting up the Enobio is extremely easy. However, the quality of recorded signals may restrict the number of classes it can use simultaneously. This study provides an evaluation of the system’s capability. Recorded data was pre-processed using EEGLAB [122]. Data was down sampled to 256Hz and re-referenced using common average. Line noise was removed with a FIR notch filter (49.5,50.5Hz). The AAR toolkit [123] was used for artefact rejection. EOG and EMG artifacts were reduced, with SOBI [198] and CCA algorithms [199], respectively. These methods outperform ICA, which is ineffective beyond 70 Hz [200, 201]. One-second epochs were extracted from the pre-processed data and saved as a numeric matrix for further analysis. Details of data acquisition and pre-processing are presented in appendix B.

## *3.6. Feature generation and subset selection*

The Discrete Gabor Transform [202, 203] was used to generate features. The original data can be reconstructed from the features with no information loss. Each Gabor coefficient contains information on both time and frequency. In this chapter, a time step of 0.015625 seconds (1-second trials, 64 steps) and a frequency band of 2Hz (0-128Hz range, 64 bands) is used for the DGT.

For one trial, using the DGT, EEG data from one channel (256 samples) is converted into a 64x64 coefficient matrix (2Hz bands 0-128Hz, and 64 time-steps in 1 second). So, for all 20 channels, **the dimension of the Gabor feature space is 1280x64.** Using this definition, the indexes of the Gabor features contain information on the EEG channel recording the data, frequency band of the feature, and time of feature. For example, Gabor feature with indexes 32x65 is linked to time 0.5 seconds, channel 2 (1 through 64 are in channel one, 65 through 128 in channel 2, and so on), frequency band 2Hz.

Instances of a Gabor feature in different trials (the respective Gabor coefficient value in that trial) are referred to as “vectors” of that Gabor feature. So, in each trial, there are 81,920 vectors (64x81920) and for each user, there are ten trials containing 819,200 vectors in total. Figure 3.4 demonstrates the Gabor feature space for one trial.

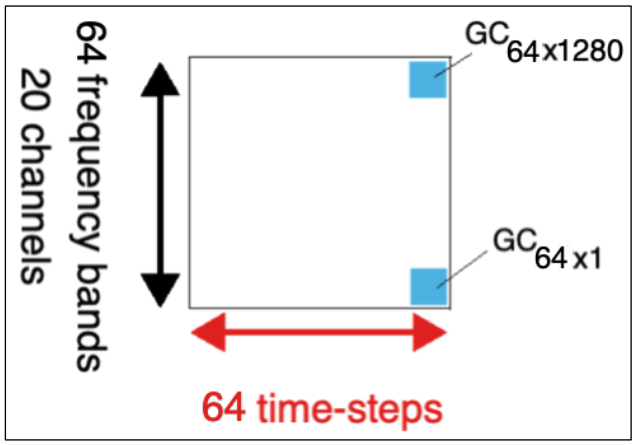



***Figure 3.4:*** *The feature space for one trial. EEG data from 20 channels for one trial is converted into 81,920 Gabor features (20 channels * 64 frequency-bands * 64 time-steps).*

Figure 3.5 presents the definition of the Discrete Gabor Transform.

$$GC_{mn} = \sum_{l=0}^{L-1} signal\ (l+1)\ e^{(-2\pi\, lm/M)}\ conj\ (g(l-an+1))$$

L = length of signal = 256 samples = 1 second

a = time step = 4 samples

M = number of frequency channels = 64

m = 0 , ... , 63 frequency index

n = 0 , ... , 63 time index

$$g(l) = \sqrt{\frac{\sqrt{2}}{T}} * e^{-\pi (l/T)}$$ Gaussian window , $T = \sqrt{a*M}$

***Figure 3.5:*** Definition of Gabor coefficients by implementation of the direct discrete Gabor transform and a Gaussian window function.

Classification true positive rate is estimated by a 5-fold cross validation process [204]. In each fold, 8 trials are used for feature selection and training the classification object, and 2 trials are set aside for testing. The most valuable Gabor features for distinguishing four classes are discovered by the Davies-Bouldin

index [205]. The process of finding two-class DBIs has been explained in section 2.6 and illustrated in figure 2.4. For calculating 4-class DBI, the process focuses on one Gabor feature at a time, completely independent of other Gabor features. In this chapter, 8 out of ten trials are used for training. Figure 3.6 shows the selection of vectors of the same Gabor feature in all four classes.

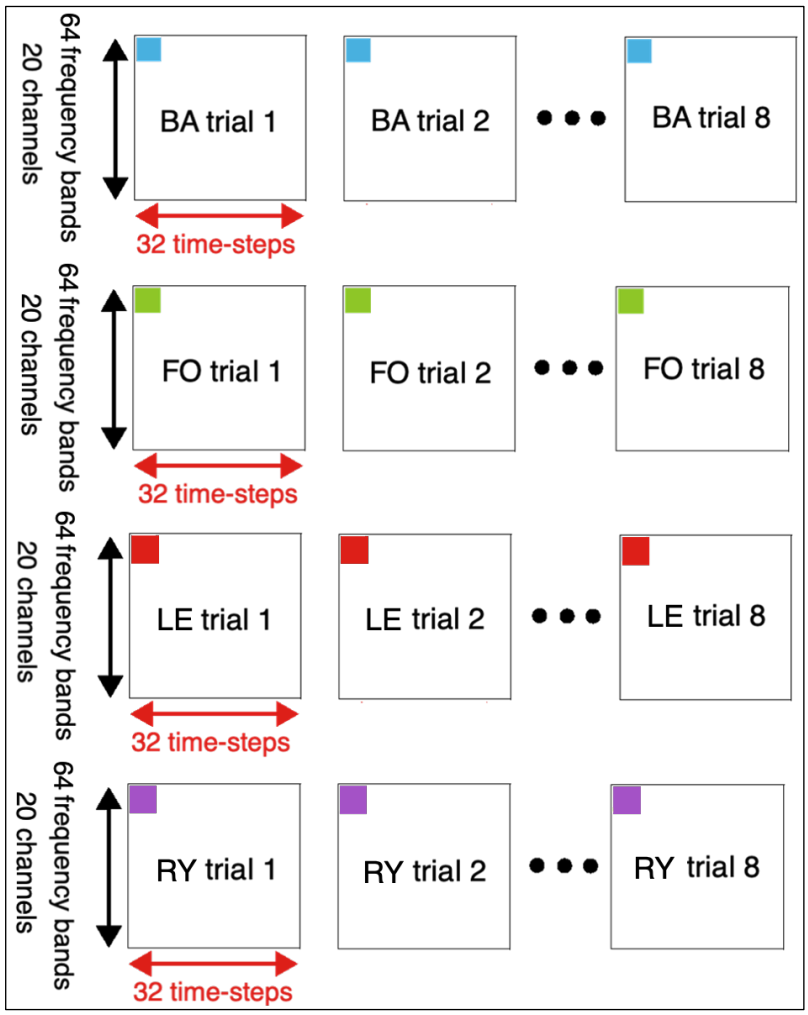



***Figure 3.6:*** *DBI is calculated for one Gabor feature at a time, independently. For example, vectors of the top right Gabor feature in all classes for all 8 training trials are used to find DBI. The curse of dimensionality does not apply here as only mean and standard deviation of 8 vectors are used for a single Gabor featurein each class.*

Initially, all pairwise DBIs are calculated (BA-FO, BA-LE, BA-RY, RY-FO, RY-LE, FO-LE). For example, *DBI(BA-FO)* is found by calculating

*(std(BA)+std(FO)) / (mean(BA)-mean(FO)),* where std(BA), and mean(BA) are the standard deviation and mean of the 8 green vectors in figure 3.6 respectively. The four-class DBI is a conservative approximation based on the two-class DBIs:

*DBI_ALL=0.25*(max([DBI_BA_FO, DBI_BA_LE, DBI_BA_RY])+*

*max([DBI_BA_FO, DBI_LE_FO, DBI_RY_FO])+*

*max([DBI_BA_LE, DBI_LE_FO, DBI_LE_RY])+*

*max([DBI_BA_RY, DBI_RY_FO, DBI_LE_RY]));*

which is defined in figure 3.7.

| | |
|---|---|
| $DBI = \frac{1}{4}\sum_{i=1}^{4} R_i$ | DB index, four classes |
| $R_i = max_{i \neq j}\left(R_{ij}\right)$ | Conservative approximation |
| $R_{ij} = \frac{S_i + S_j}{D_{ij}}$ | DB index, two classes |
| $D_{ij} = \sqrt{\left(c_i - c_j\right)^2}$ | Euclidean distance of centroids |
| $S_i = \frac{1}{N}\sum_{n=1}^{N}(x_n - c_i)$ | Standard deviation of class "i" |

***Figure 3.7:*** Definition of the Davies-Bouldin index for 4 classes. The most valuable features have the smallest DBI.

Like figure 2.5, Gabor features are sorted and ranked based on their DBI, the 3K most valuable Gabor features (from a total of 81,920 per trial) are identified. Vectors of these Gabor features (Gabor coefficients located at the same indexes), within the training trials (8 out of 10 trials in each classification fold) are used to train the Pseudo-LDA classifier. Vectors of the same Gabor features

within the testing portion of the data (2 out of 10 trials per classification fold) are used to test the performance of the PLDA classifier. The PLDA has been described in detail in section 2.7 in the previous chapter. The 4-class DBI,which is based on binary DBI is also immune from the curse of dimensionality.

## *3.7. Classification accuracy and analysis of results*

The true positive rates for one class vs. all, are estimated as the mean and standard deviation of the five-fold cross validation process. Table 3.1 contains these results for the four participants and both types of cognitive task. The reader should kindly consider that the objective of this study is not to maximize classification accuracy. The experimental protocol and analysis pipeline provided identical environments for both paradigms, while identifying the most important activities related to the selected features.

| | *Covert Speech* | *Motor Imagery* |
|---|---|---|
| *User 1* | 85 ± 33.3 | 80.1 ± 32.7 |
| *User 2* | 80.5 ± 30.8 | 68.5 ± 28.1 |
| *User 3* | 87.3 ± 21.2 | 83.4 ± 33.4 |
| *User 4* | 78 ± 18.9 | 78 ± 30.9 |
| *Average* | 82.5 ± 24.1 | 77.2 ± 31.2 |

***Table 3.1.*** True positive rates of one class vs. all. These are estimated using a five-fold cross validation process. With four classes, the average performance is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect.

To demonstrate feature separability of covert speech tasks, vectors of the Gabor features (Gabor coefficients) in one trial are plotted for each of the four classes, and presented in figures 3.8, 3.9, 3.10, and 3.11 for users 1, 2, 3, and 4 respectively. The vertical axis has 64 time-steps, and the horizontal axis has 1280 points from (20 channels with 64 bands per channel). There are visible differences in the patterns.

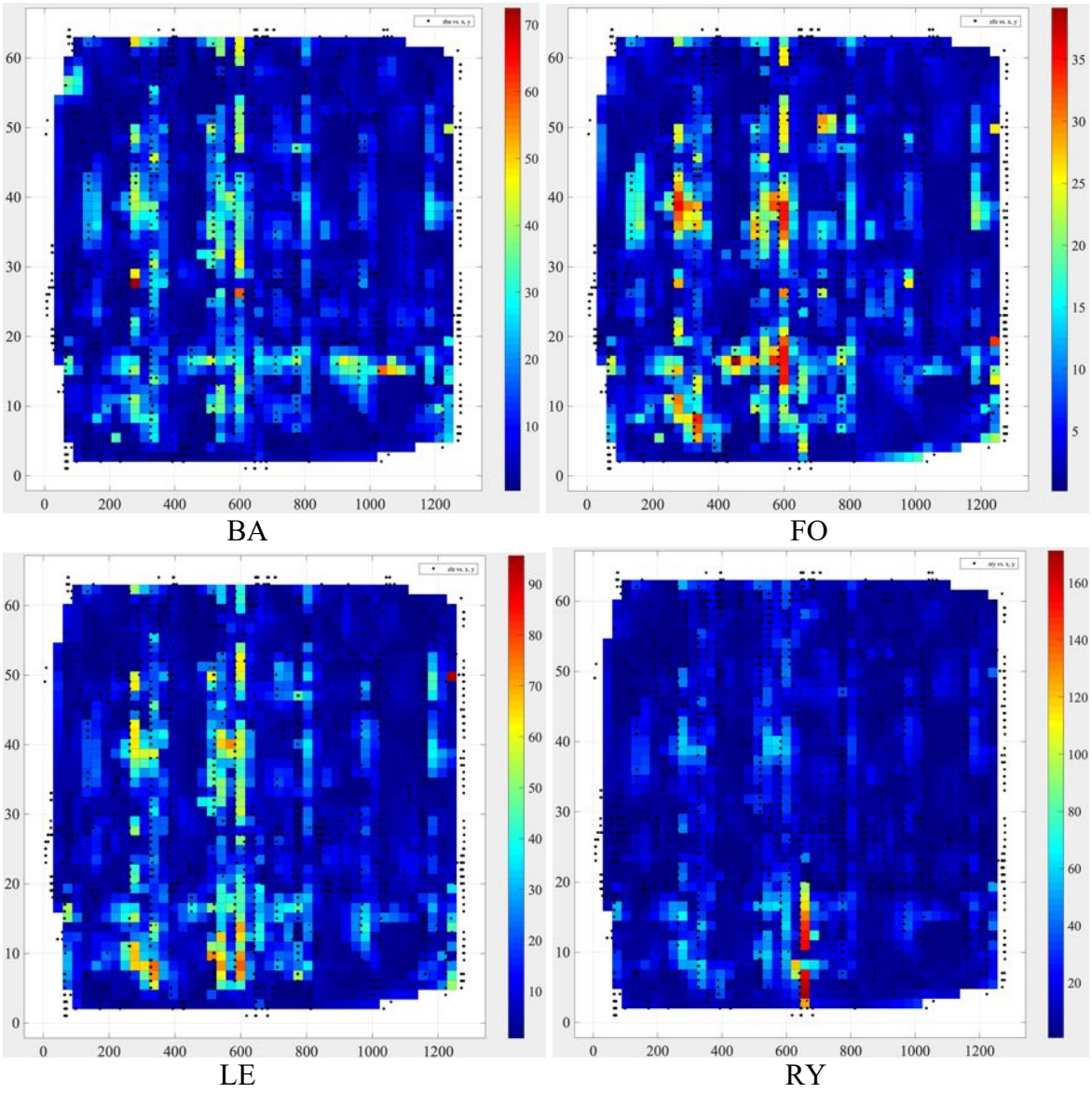



***Figure 3.8:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 1. The vertical axis has 64 time-steps, and the horizontal axis has 1280 points from 20 channels with 64 frequency bands per channel.*

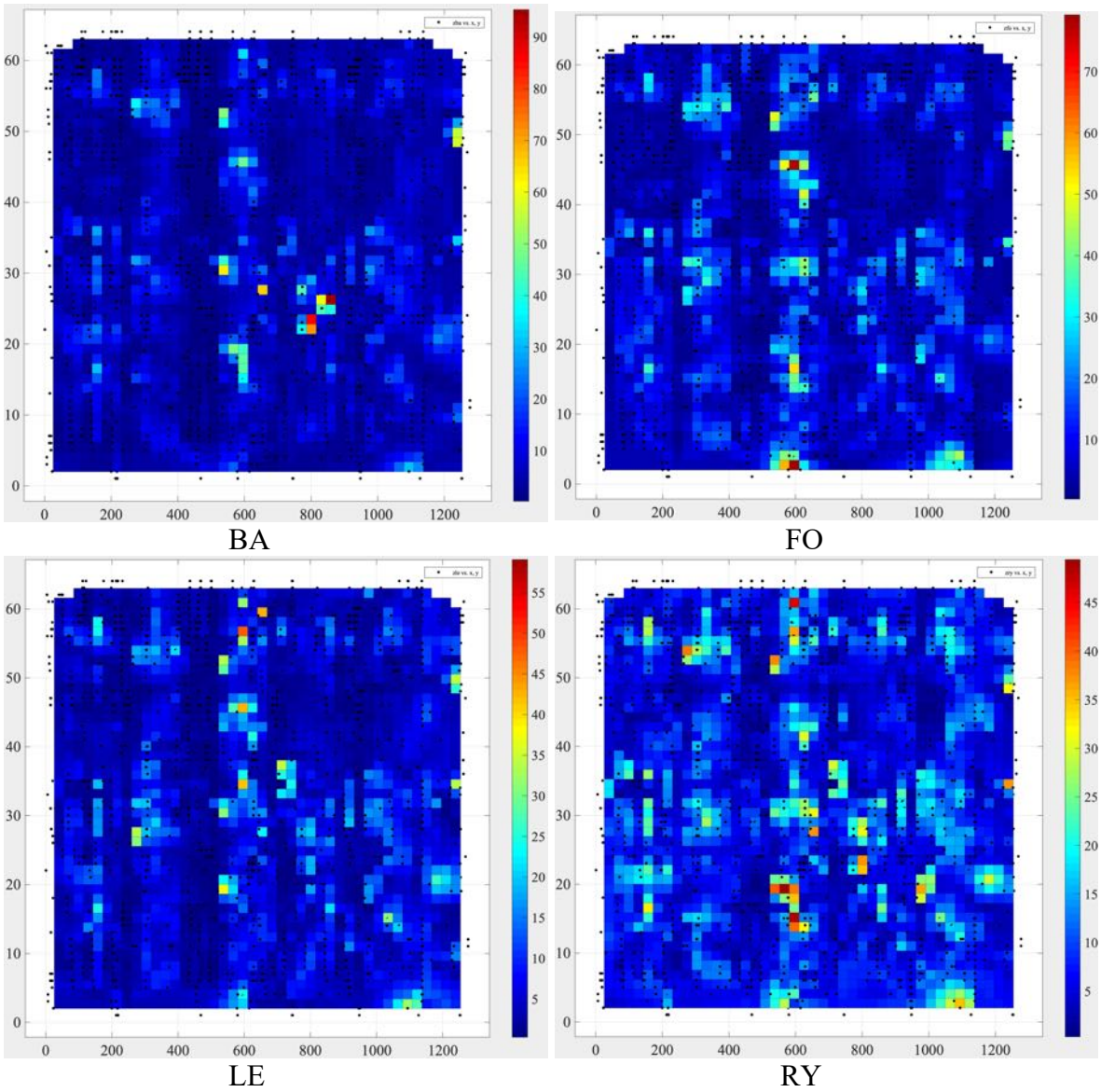



***Figure 3.9:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 2. The vertical axis has 64 time-steps, and the horizontal axis has 1280 points from 20 channels with 64 frequency bands per channel.*

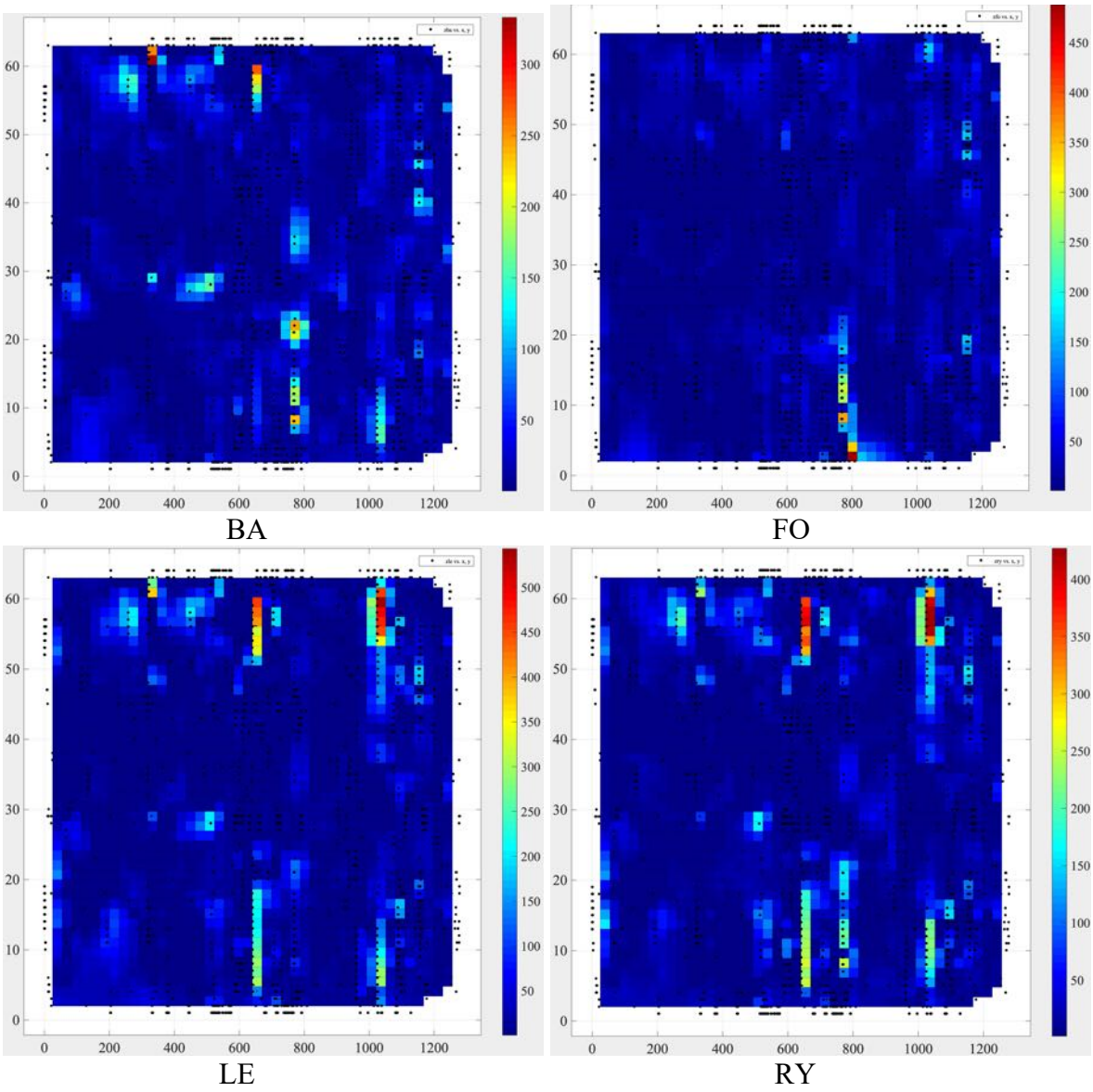



***Figure 3.10:*** *Vectors pf the Gabor features from covert speech of one trial, recorded from user 3. The vertical axis has 64 time-steps, and the horizontal axis has 1280 points from 20 channels with 64 frequency bands per channel.*

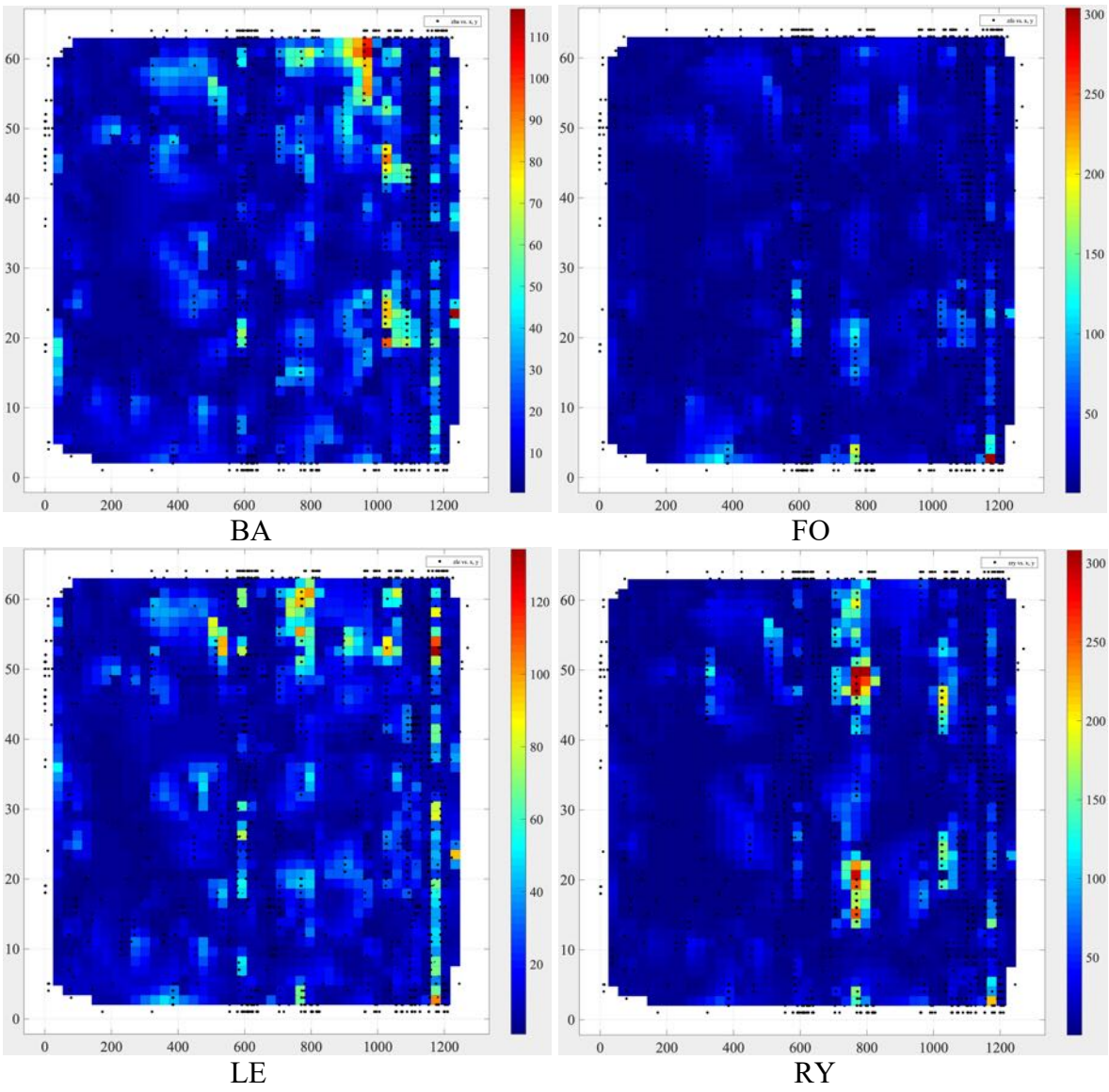



***Figure 3.11:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 4. The vertical axis has 64 time-steps, and the horizontal axis has 1280 points from 20 channels with 64 frequency bands per channel.*

With four classes, the classification accuracy is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect. The outstanding times and frequencies, based on the number of times they have been identified in the most valuable Gabor features, in all validation folds, for all users are sown in figure 3.12. As expected, valuable

class dependent activity is not limited to the Alpha and Beta bands. In addition, the nominal bandwidth of (1-125) Hz given by Enobio is confirmed, as valuable vectors are identified in the entire frequency range.

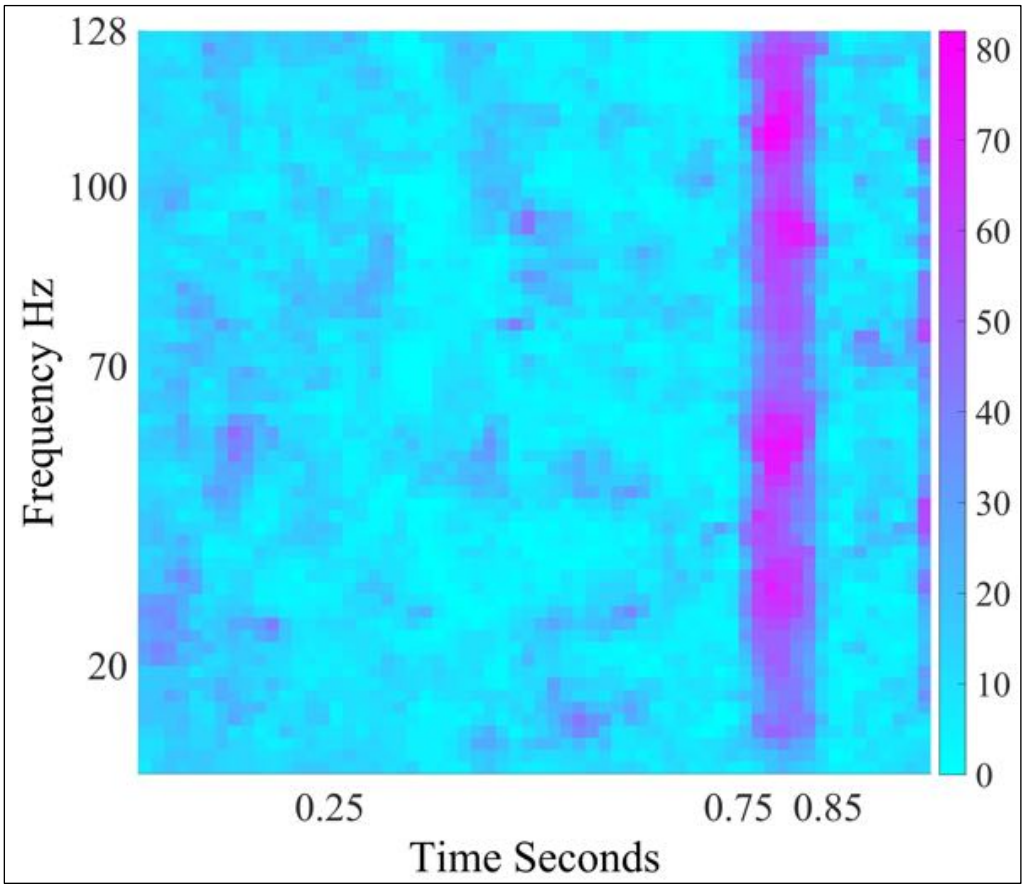



***Figure 3.12:*** *The cumulative joint time-frequency plot containing the most valuable 60K features identified in the motor imagery experiments (4 users, five validation folds, and 3K vectors per fold). The (0.73,0.875) second band contains 15.1% of vectors.*

For motor imagery tasks in this experiment 15.1% of all the most valuable vectors are significantly concentrated within the (0.73-0.875) second range. This time period corresponds with performing imagined movements and the suppression of the Primary Motor Cortex (stopping actual movements) via "goal

driven executive control". Such executive control involves high-frequency cognitive activity in brain regions such as the Superior Parietal Cortex and the Pre-Frontal Cortex [160, 161]. 23.2% of all the most valuable vectors are within the Alpha and Beta bands (MI). The other 76.8% of the vectors are in the Gamma, and high-Gamma bands (cognitive functions). This suggests that in motor imagery tasks, cognitive functions generate a significantly greater amount of class dependent activity compared to the execution movement. In Figure 3.13 the outstanding times and frequencies, based on the number of times they have been identified in the most valuable Gabor features, in all validation folds, for all users (3K/fold, 4 users, 5 folds/user=60K) are shown for covert speech experiments.

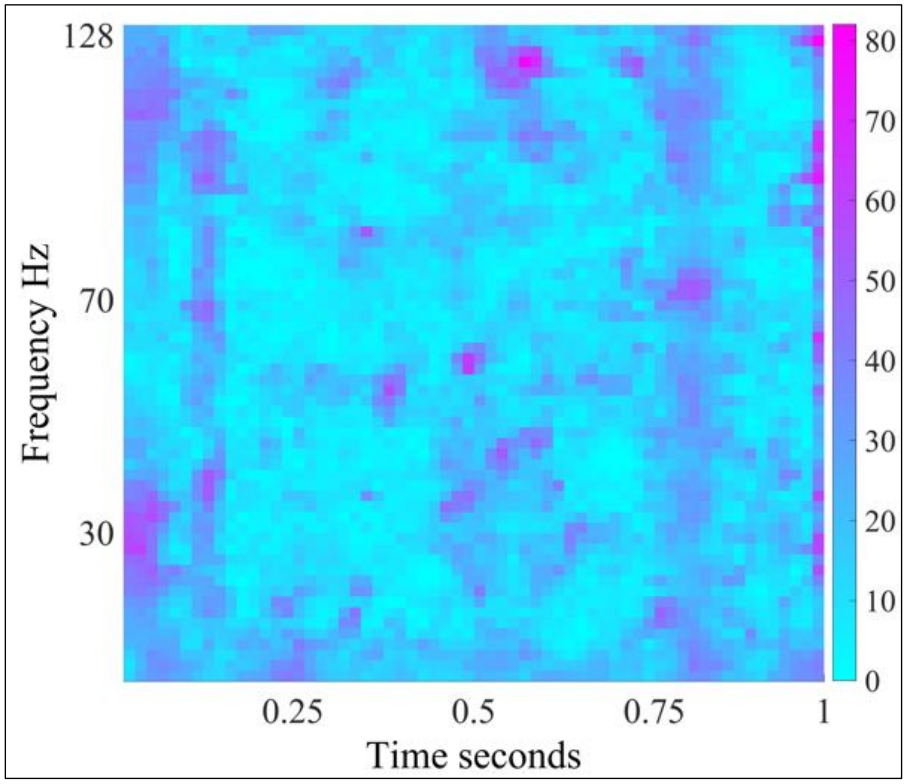



***Figure 3.13:*** *The cumulative joint time-frequency plot containing the most valuable 60K Gabor features identified in the covert speech experiments (4 users, five validation folds, and 3K features per fold). 48.8% of these are above 70 Hz. The (0.73,0.875sec) band is not as prominent as the MI paradigm from figure 3.12.*

48.8% of these vectors are above 70 Hz, which correspond with the linguistic processing functions [48]. These linguistic functions, which are entirely class dependent, do not exist in motor imagery. This provides a possible explanation for the higher classification accuracy of covert speech tasks (82.5%) compared to motor imagery tasks (77.2%) in an identical environment, considering there is a direct positive correlation (with R=0.8822 and P=0) between their performances. Considering that tasks are identified before trials begin, the cognitively demanding linguistic functions (conceptual preparation, Lemma selection) are completed before onset. The linguistic functions occurring within trials (phonological code retrieval, syllabification) are performed automatically by the brain [27] and require no user effort. All other cognitive functions within trials (executive control, imagined movement) are also present in MI tasks. As a result, the cognitive effort of using covert speech tasks and MI tasks are virtually identical in this study.

## *3.8. Discussion*

The linguistic processing stages of word production prior to articulation, which are entirely class-dependent, consist of conceptual preparation, Lemma selection, phonological code retrieval, and syllabification [39]. By incorporating difference in meaning, and difference in phonetic structure, for selecting covert speech tasks, class separability can be significantly enhanced. In this experiment,

linguistic class separability is maximized by selecting phonetically dissimilar covert speech classes [196]. This explains the superior performance of covert speech tasks compared to MI tasks in the otherwise identical environment designed in this study.

Linguistic studies using intra-cranial implants have demonstrated that these linguistic processing stages have high-Gamma signatures in the (70-170Hz) range [23-25, 34, 206]. As bandwidth of EEG systems increases and EMG removal algorithms become more reliable, covert speech BCIs will become much more capable. Although other BCI systems (such as MI) will also improve, language, which is the most intuitive and natural form of human communication, would logically be the preferred paradigm of choice for a BCI. In this chapter we used mixed randomized recording sessions containing all four classes.

In order to test the hypothesis that classification accuracy of the covert speech tasks has been achieved by chance, accuracies obtained on the baseline resting state data, and on data with random label permutations are calculated using the same analysis pipeline. This allows the evaluation of the obtained actual class data classification accuracy against the chance level accuracy. As seen in table 3.2, the rest state accuracy from 1 second pre-cue data, and that of data with randomized labels are close to chance level. Thus, the above hypothesis is rejected.

| | *Covert speech 1s epoch* | *1 second pre-cue* | *Randomised Labels* |
|---|---|---|---|
| *User 1* | *85 ± 33.3* | *27.5±20.5* | *20±18.9* |
| *User 2* | *80.5 ± 30.8* | *25±8.83* | *17.5±14.2* |
| *User 3* | *87.3 ± 21.2* | *17.5±6.8* | *17.5±11.1* |
| *User 4* | *78 ± 18.9* | *20±6.9* | *20±11.1* |
| *Average* | *82.5 ± 24.1* | *22.5±10.75* | *18.75±13.8* |

***Table 3.2:*** *classification accuracy for four covert speech tasks "BA", "FO", "LE", and "RY" are presented in the first column.. 1-sec pre-cue (Idle) performance, and randomised labels performance are below chance level.*

In addition, to demonstrate that the only time range in the data with high feature seperability is only within the task duration (0-1sec post-onset), classification accuracy for trials 1-2 sec, 2-3 sec, 3-4 sec, and 4-5 sec post-onset, are calculated. As seen in table 3.3, the classification accuracy for all these rest-state trials is close to chance level.

| | *1 to 2 sec* | *2 to 3 sec* | *3 to 4 sec* | *4 to 5 sec* |
|---|---|---|---|---|
| *User 1* | *30±18.9* | *22.5±10.4* | *35±5.5* | *29±11.1* |
| *User 2* | *20±6.8* | *12.5±12.5* | *20±14.2* | *25±8.8* |
| *User 3* | *17.5±11.1* | *22.6±16.2* | *10±10.4* | *17.5±6.8* |
| *User 4* | *20±18.9* | *35±10.4* | *22.5±20.5* | *27.5±10.4* |
| *Average* | *21.9±13.9* | *23.2±12.4* | *21.8±12.7* | *24.8±9.2* |

***Table 3.3:*** *The time course of accuracy after cue, excluding the first trial containing Linguistic activity.*

Finally, to demonstrate that the Pseudo-LDA classifier is not affected by the curse of dimensionality when 3K Gabor features are used, the performance for fewer Gabor features are calculated. As seem in table 3.4, with the increase in the number of Gabor features, performance increases consistently and steadily.

| | *20 features* | *100 features* | *300 features* | *1000 features* | *2000 features* | *3000 features* |
|---|---|---|---|---|---|---|
| *User 1* | *67.5±18.9* | *57.5±14.2* | *60±22.3* | *77.5±29.8* | *82.5±32.5* | *85 ± 33.3* |
| *User 2* | *47.5±18.5* | *77±24* | *80±32.5* | *82.5±25.9* | *80±31.3* | *80.5 ± 30.8* |
| *User 3* | *40±10.4* | *57.5±14.2* | *67.5±16.7* | *87.5±15* | *90±16.2* | *87.3 ± 21.2* |
| *User 4* | *57.5±16.7* | *60±5.5* | *72.5±13.6* | *75±19.7* | *75±15.3* | *78 ± 18.9* |
| *Average* | *53.1±16.1* | *63±14.5* | *70±21.3* | *80.6±22.6* | *81.9±23.8* | *82.5 ± 24.1* |

***Table 3.4:*** *Predicted performance with 3000 vectors vs. performance with less vectors using Pseudo Linear Discriminant Analysis. As more vector are used, performance increases steadily and consistently.*

Table 3.4 will us to compare the estimated performance and the real-world performance of the BCI in chapter five, to see if there is overfitting, and if so, to what extent this is true.

Finally, to completely reject the Null Hypothesis, for each user, within each validation fold, 100 randomised lable tests are performed. the average of these 100 randomised tests are compare with the correct lable test. These are shown in table 3.5. The grand average spectrogram for each user for each class in the central channel Cz are presented in figures 3.14 to 3.17. There are clear differences in the plots for each class. Although feature power is not used for classification and the lower frequency bands are much more visible in the plots, there are still clear differences in the plots for each class.

| | | *Fold 1* | *Fold 2* | *Fold 3* | *Fold 4* | *Fold 5* | |
|---|---|---|---|---|---|---|---|
| *User 1* | *Correct* | *100* | *87.5* | *100* | *87.5* | *50* | *P-value=0.0079* |
| | *Random* | *21.5* | *19.38* | *20.63* | *21.75* | *16.75* | *h=1* |
| *User 2* | *Correct* | *87.5* | *87.5* | *75* | *100* | *50* | *P-value=0.0081* |
| | *Random* | *17.5* | *18.13* | *17.25* | *17.62* | *17.12* | *h=1* |
| *User 3* | *Correct* | *100* | *100* | *87.5* | *100* | *50* | *P-value=0.0082* |
| | *Random* | *18.37* | *17.25* | *17.5* | *17.87* | *16.62* | *h=1* |
| *User 4* | *Correct* | *100* | *87.5* | *75* | *75* | *50* | *P-value=0.008* |
| | *Random* | *21.5* | *18.75* | *19.62* | *20.62* | *19.5* | *h=1* |

***Table 3.5:*** *Rejecting Null Hypothesis: For each user, within each of the 5 cross validation folds, the performance is tested with randomised labels 100 times. The average of these 100 tests with randomised labels is shown under the performance with correct labels for each cross-validation fold. Wilcoxon rank-sum tests are performed. The value of h=1 rejects the null hypothesis suggesting the results are indeed statistically significant and have not been achieved by chance.*

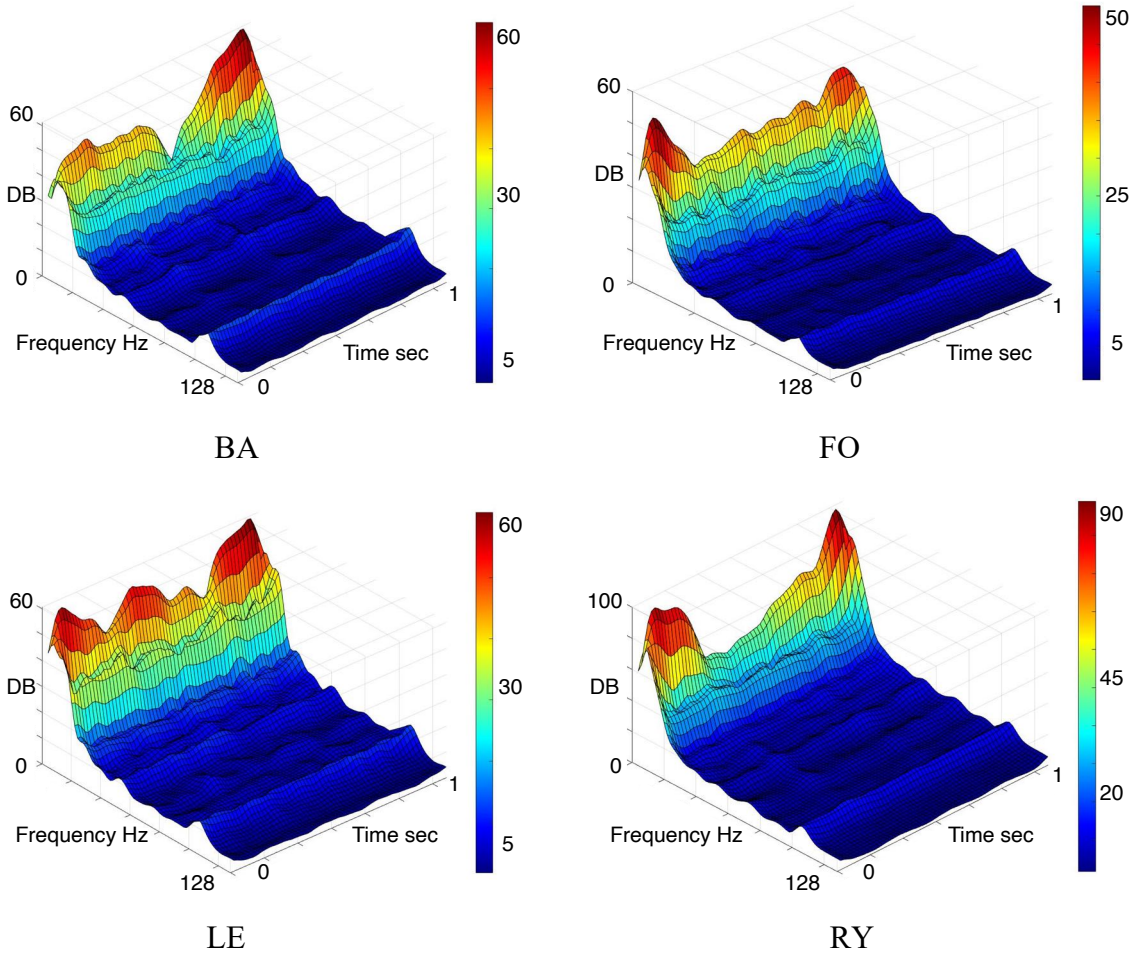



***Figure 3.14:*** *User 1: Grand average spectrogram of each class from all 10 trials and all channels.*

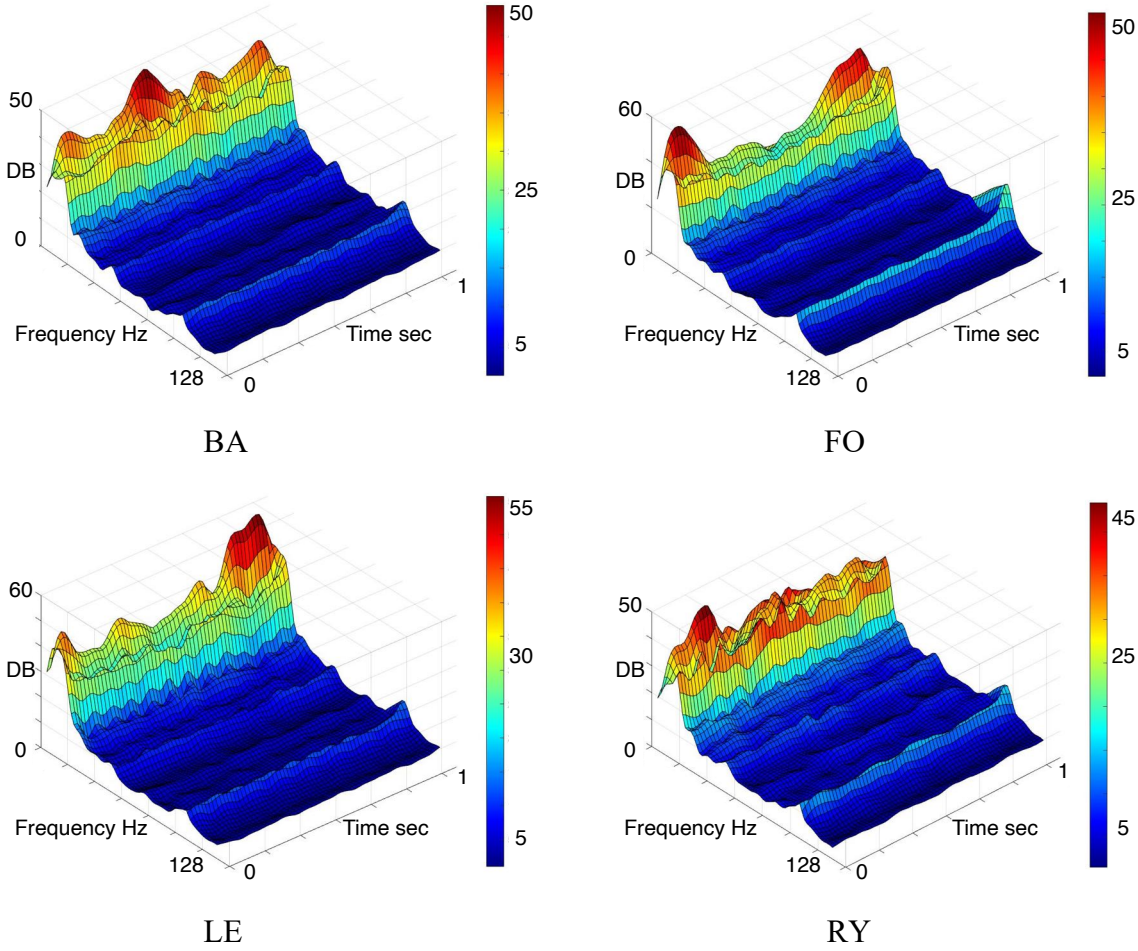



***Figure 3.15:*** *User 2: Grand average spectrogram of each class from all 10 trials and all channels.*

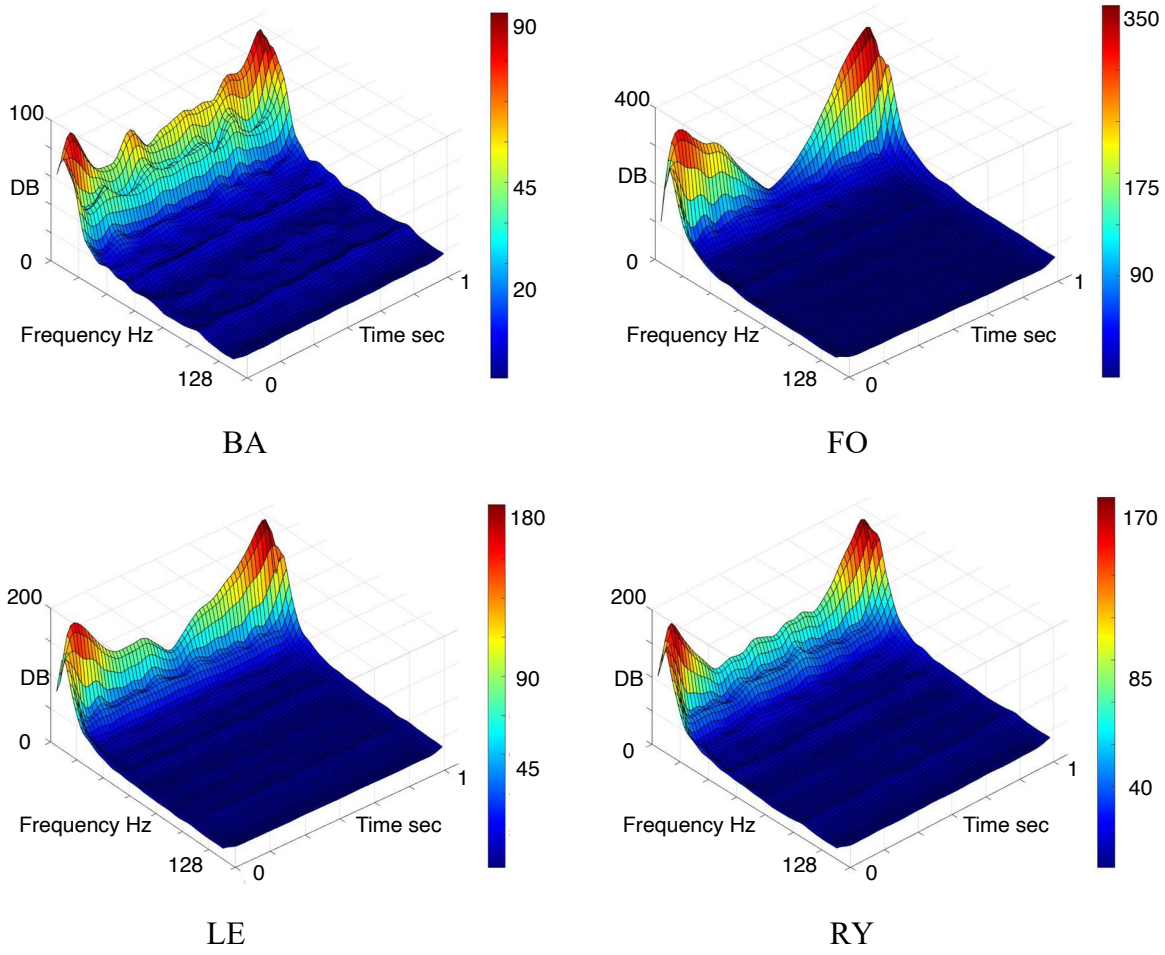



***Figure 3.16:*** *User 3: Grand average spectrogram of each class from all 10 trials and all channels.*

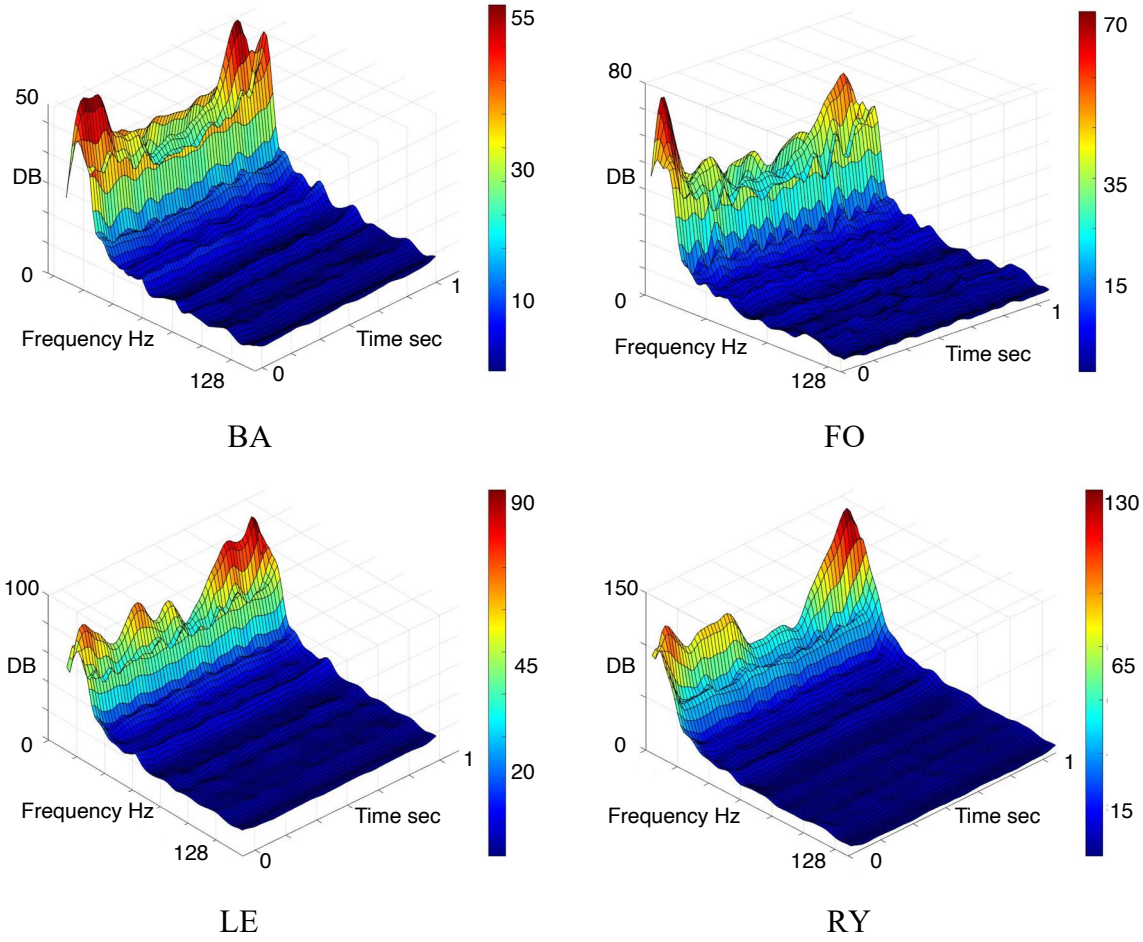



***Figure 3.17:*** *User 4: Grand average spectrogram of each class from all 10 trials and all channels.*

# CHAPTER FOUR:
## OPTIMIZING THE SYSTEM

## *4.1. Refining the experimental protocol*

In chapter three, the experimental protocol was designed to show the user the next task in advance. As a result, the conceptual preparation and lemma selection stages of linguistic processing is completed. The, a constant stand-by period of 3 seconds was used. The user could guess the time of activation cue in advance, and as a result, the phonological code retrieval and syllabification stages of linguistic processing would be completed before the start of trials. The only activity recorded in these trials would be the motor imagery of articulation. In this chapter, the stand-by period after representing the next task, and prior to the onset cue, is a random duration of 3-7 seconds. **This is critically important**.

As a result, after conceptual preparation and lemma selection, the user waits for cue recognition, to perform the remaining activities. Such trials would contain signatures of cue recognition, and the remaining linguistic processing stages, which are phonological code retrieval, syllabification, and planning movements for motor imagery of articulation. 1-second trials would contain all these stages. Auditory cue recognition, phonological code retrieval, and syllabification are complete within 300ms. By using trials with 312ms duration, only linguistic activity is recorded, and trials are complete before motor imagery of articulation has even begun. By comparing 1-second trials as a baseline, to 312ms trials, the contribution of motor imagery of articulation to classification

compared to linguistic activity alone, is studied. Figure 4.1 presents the refined experimental protocol.

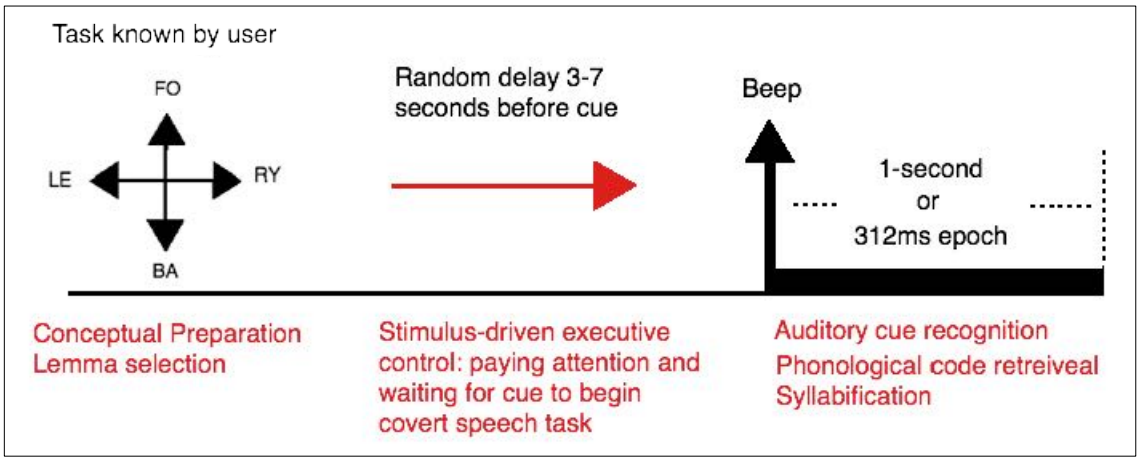



***Figure 6.1.*** *Imagination protocol. The user imagines speaking a word when an auditory cue in the form of a beep is presented. One second after each cue are used for the first experiment, and 312ms for the second. A random rest period of 3-7 seconds occurs between trials. This sufficiently separated the tasks from one another. Also, the random duration prevents the user from anticipating the task onset based on rhythm. As a result, the next stages of linguistic functions begin exactly after cue recognition and the system is perfectly synchronized.*

## *4.2. Methods*

This study was conducted with 10 neurologically healthy volunteers in the age group of 21-33. All volunteers signed a consent form based on the recommendations of the Ethical Committee of the University of Essex. Participants were seated in a comfortable armchair. 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 EEG signals were recorded using a 64 channel Biosemi ActiveTwo$^{TM}$ system [207]. One computer generated the graphical user interface and sent

trigger signals to the ActiveTwo device at the instant a time cue was presented to the user. The triggers were sent via the parallel port and were visible in the recorded data. A second computer saved the EEG recordings and was connected to the ActiveTwo's A/D box via USB. Electrode placement was done per the international ABC system, which for 64 channels corresponds to the 10/10 system. The ActiveTwo has a pre-amplifier stage on the electrode and can correct for high impedances. However, the offset voltage between the A/D box and the body was kept between 25mV and 50mV as recommended by the manufacturer. The data were recorded at a sampling rate of 2048 samples/s, with guaranteed data frequency content of 0-409Hz according to BioSemi.

The pre-processing was done with the use of EEGLAB [122], Similar to the previous chapter, and the description in appendix A. The final stage of pre-processing is extracting epochs from the continuous EEG recordings. Each epoch begins when beep sound is generated and ends exactly one second (or 312ms for shortened trials) later.

The Discrete Gabor transform is used for feature extraction in this chapter also. A time-step of 0.03125 seconds (32 steps per second) and frequency band of 2Hz (64 frequency bands) were chosen. A 1-second epoch from a single EEG channel (256 samples) is converted into a 64x32 matrix. For the 312ms trials (80 samples), one epoch from one channel is converted into a 64x10 matrix.

For the 312ms trials, there are 10 time-steps, 64 channels, and 64 frequency bans, and the Gabor feature space for one trial has the dimension of 10x4096 as seen in figure 4.2. There are 30 trials, and a total of 1,220,700 vectors of these Gabor feature (40960 vectors per trial x 30 trials).

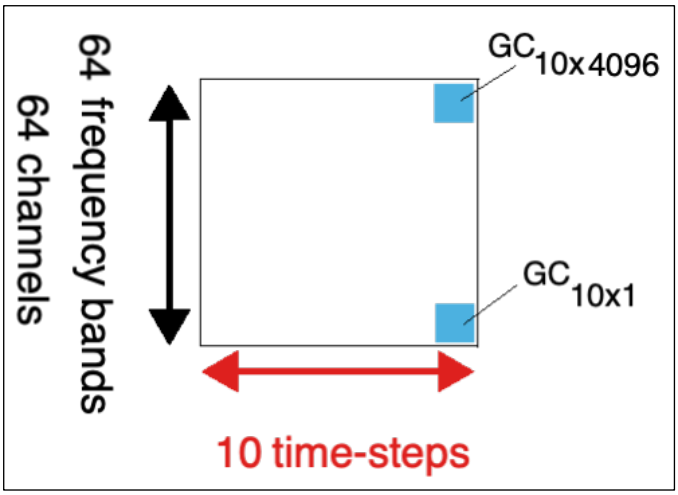



***Figure 4.4:*** *The Gabor feature space for the 312ms trials. There are 10 time-steps, and 64 channels containing 64 frequency bands each.*

Feature selection using DBI is performed in a similar manner to the previous chapter. The only difference is that the most valuable 4K Gabor features are selected. Similarly, the DBI for each Gabor feature is found by using vectors of that Gabor feature in the training data (27 out of 30 trials this chapter) for all four classes.

The mean and standard deviation of a 10-fold cross validation process [204] were used to estimate the true positive rate. For each validation fold, 27 trials were used for training, and 3 remaining trials were set aside for testing only. Testing trials change from one validation fold to the next, and over 10 folds, all

30 trials are used at least once in testing. The process of cross validation, feature selection, training, and testing used in this work is presented in figure 4.5.

Pseudo-Linear discriminant analysis was applied for classification, as it consistently out-performed all other supervised machine learning methods, for EEG recorded covert speech data [208]. Compared to the training process, the computational cost of testing is negligible.

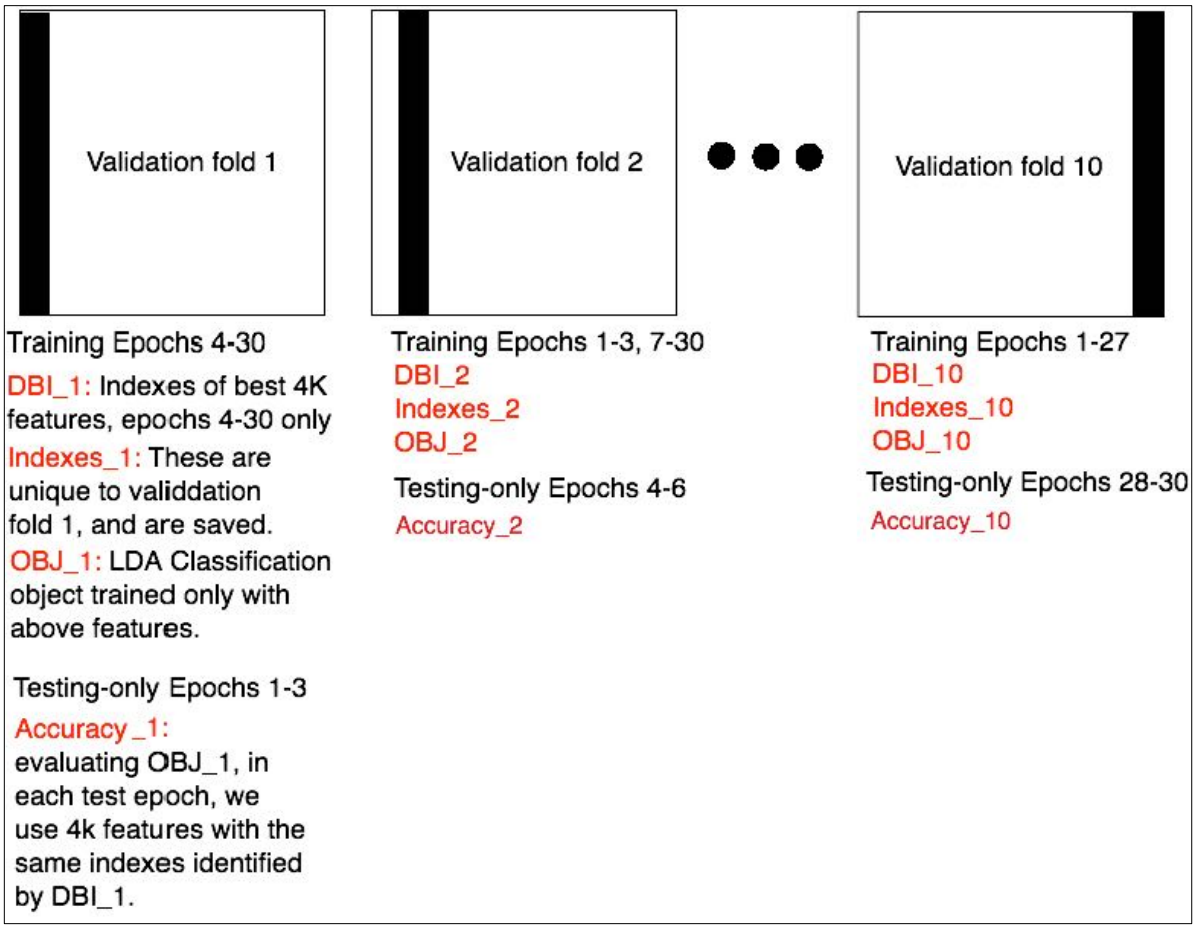



***Figure 4.5*** The process of cross validation, feature selection, training, and testing used in this work is presented here. The grand average true positive rate is the mean and standard deviation of "Accuracy_1" through "Accuracy_10".

## *4.3 Results*

The true positive rates of one word vs. all classified by PLDA, are generated by a standard ten-fold cross validation method. Table 4.1 presents these values for 1-second epochs, and for 312ms epochs. By eliminating the covert articulation stage from trials, the relative contribution of Motor Imagery of speech and linguistic processing stages, in classification accuracy can be determined.

| | 1-second epoch | 312ms epoch |
|---|---|---|
| U1 | 95.8±9% | 96.7±7% |
| U2 | 95±10.5% | 94.1±8.8% |
| U3 | 100% | 100% |
| U4 | 98.3±5.2% | 97.5±5.6% |
| U5 | 94.1±7.9% | 92.5±10% |
| U6 | 93.2±10.9% | 89.1±9.6% |
| U7 | 96.6±8% | 91.6±11.1% |
| U8 | 94.1±7.9% | 89.1±8.8% |
| U9 | 99.1±2.6% | 100% |
| U10 | 97.5±5.6% | 94.1±7.9% |
| Mean | **96.4±2.3%** | **94.5±4%** |

***Table 4.1:*** *The true positive rates of one word vs. all, estimated by a ten-fold cross validation method. Eliminating the covert articulation stage from analysis has less than 2% effect on grand average classification accuracy. Considering the Wilcoxon p-value of 0.9269, compared to the high-Gamma linguistic processing stages, the contribution of motor imagery of articulation in class separation of covert speech tasks from EEG data is negligible.*

The Wilcoxon rank-sum test on both columns returns a p-value of 0.9269. By using 312ms trials instead of 1-second trials to exclude covert articulation, the

computational cost is reduced to one third, with less than 2% penalty in classification accuracy. During covert speech, the language motor regions are suppressed, but not completely deactivated [48]. As a result, during the covert articulation stage, there will be minute involuntary muscle movements related to each phonemic structure, which will create class-related, high-Gamma Myoelectric artefacts. The 312ms trials are complete before the covert articulation stage begins (~500ms post onset) and are guaranteed to be free from class-related EMG. Possible involuntary early muscle ticks (i.e. lip movements ~160ms after cue) can cause significant EMG contamination. The CCA algorithm used here, only removes such artefacts from the first 400ms of data (312ms trials included) [209].

As seen in table 4.1, the contribution of motor imagery of articulation, compare to that of the linguistic processing stage, in classification accuracy is negligible. So, the experimental settings and analysis pipeline for the 312ms trials are chosen as the final design option for use in an online system in the next chapter. To demonstrate feature separability of covert speech tasks, the vectors of Gabor features (Gabor coefficients) from one trial, and for all four classes are presented in figures 4.6 to 4.15, for users 1 to 10 respectively. There are visible differences in the patterns of the vectors for each word class.

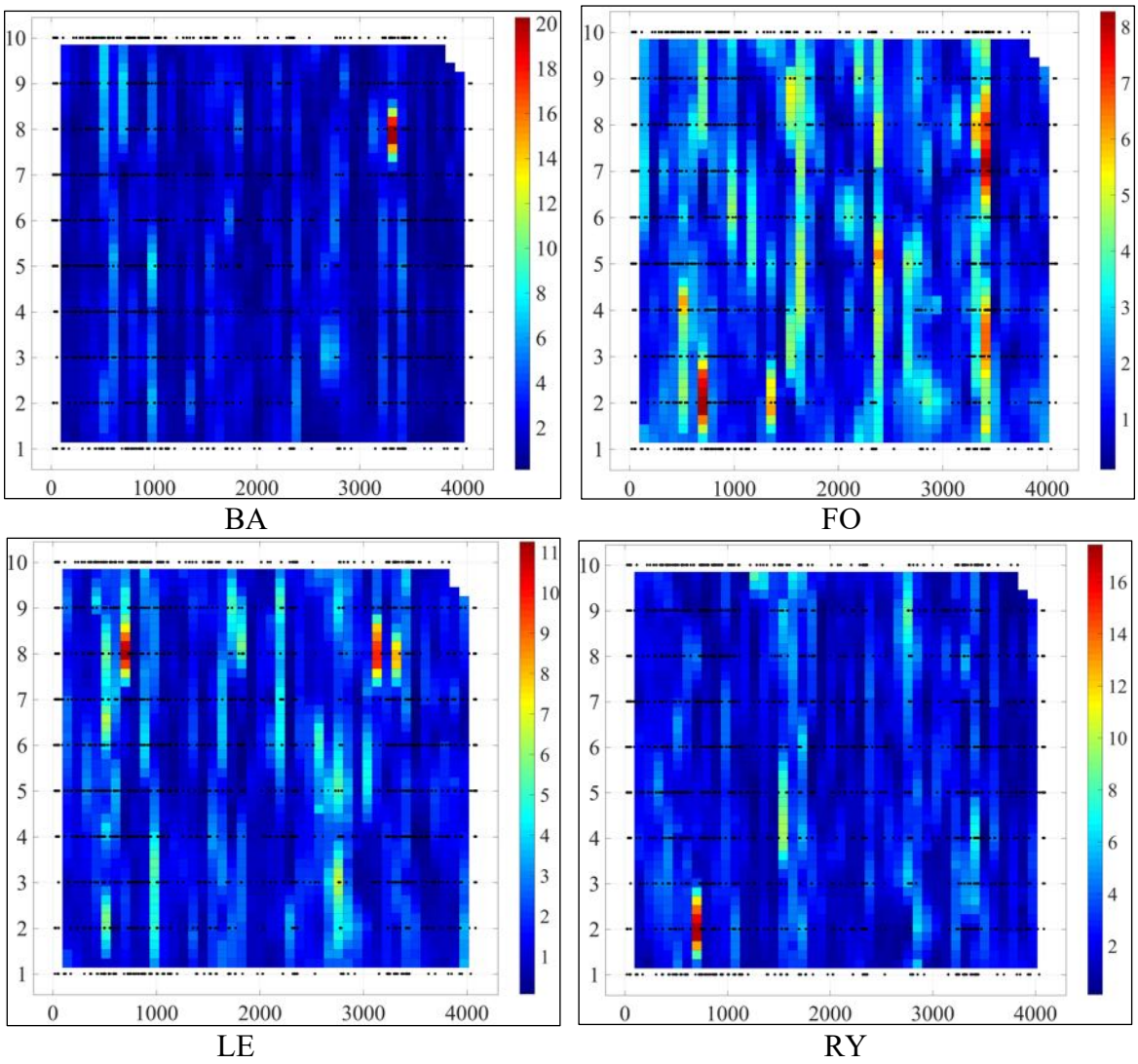



***Figure 4.6:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 1. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

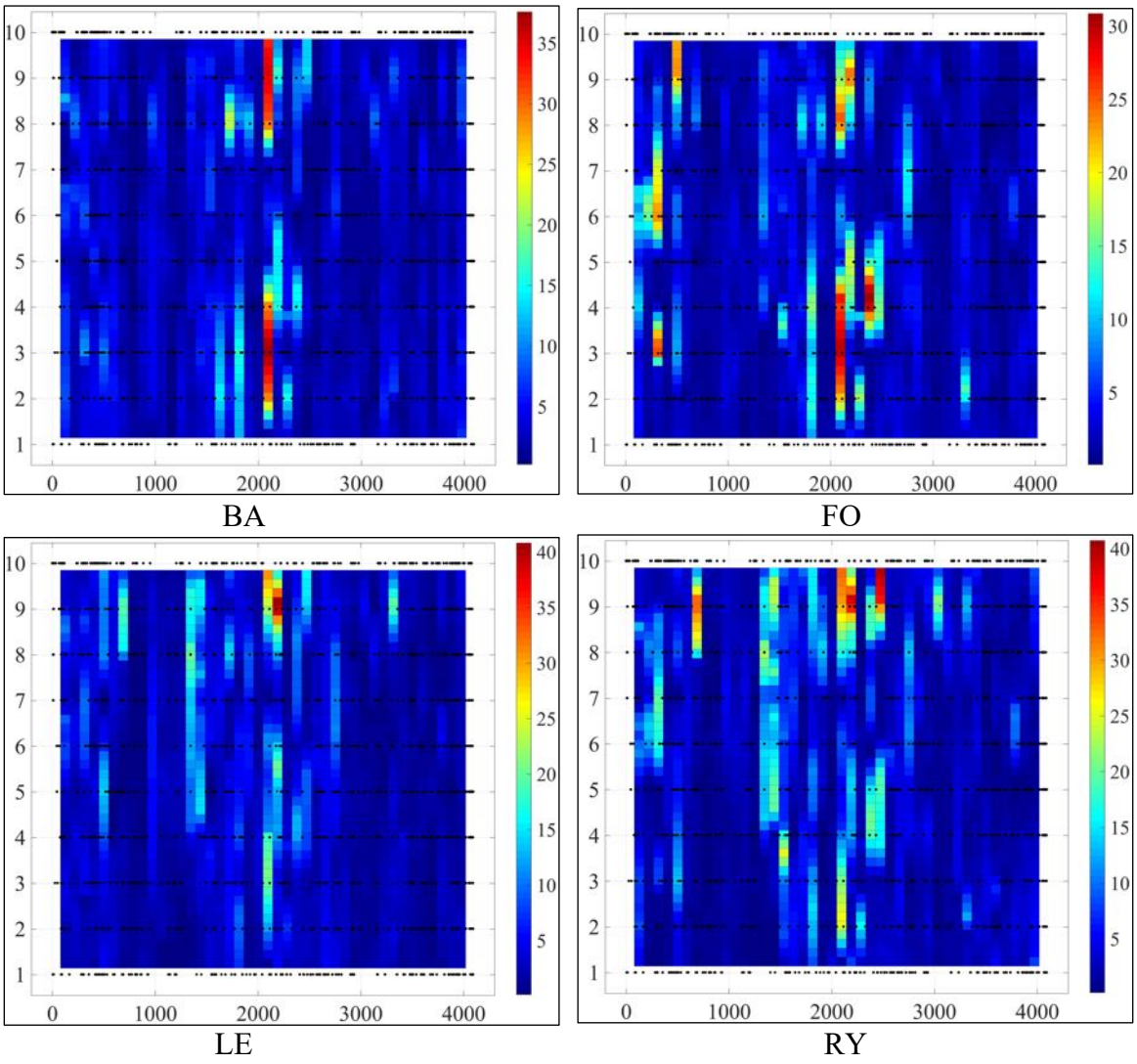



***Figure 4.7:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 2. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

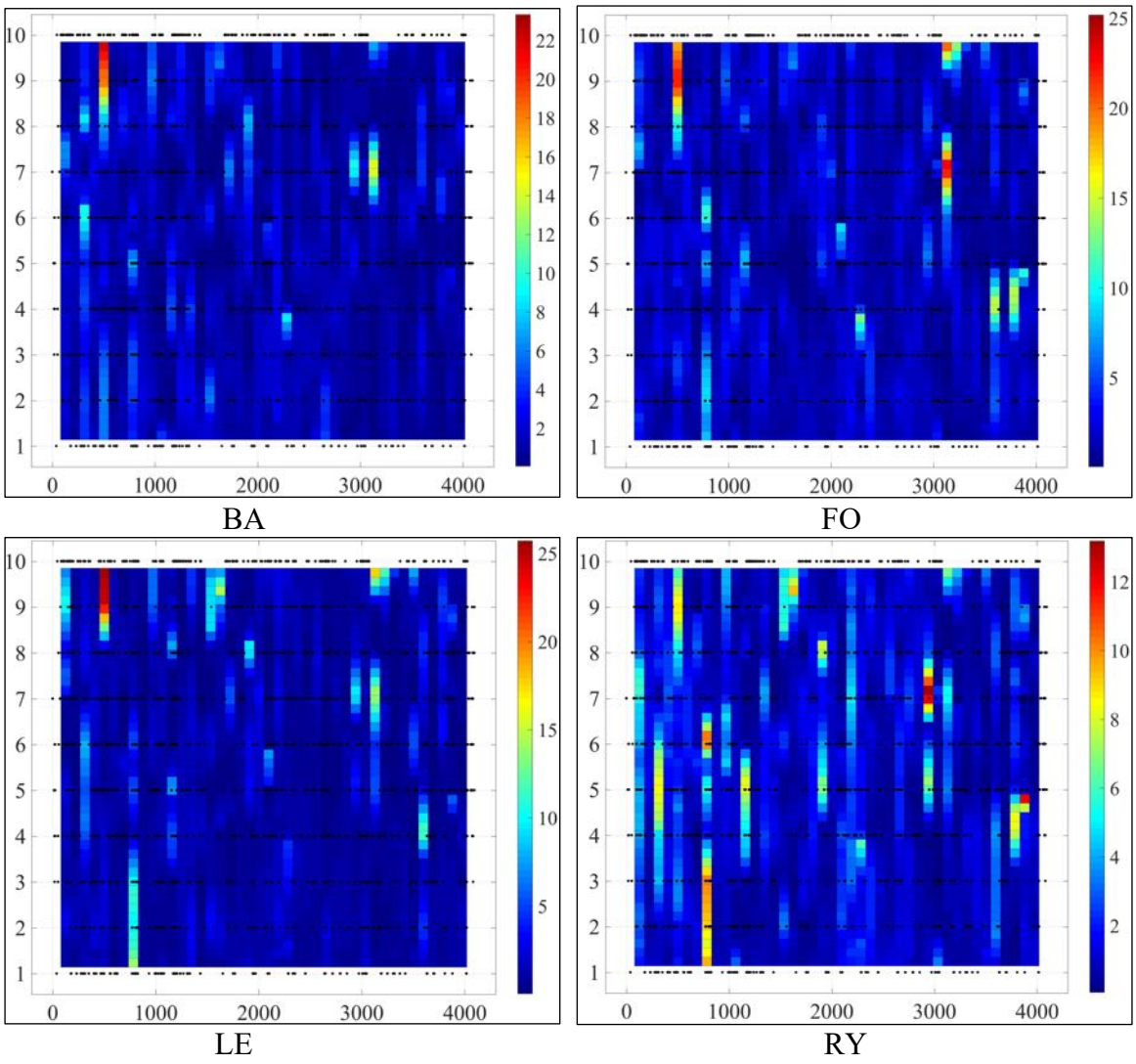



***Figure 4.8:*** *Vectors of the Gabor features from covert speech of one trial, recorded from user 3. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

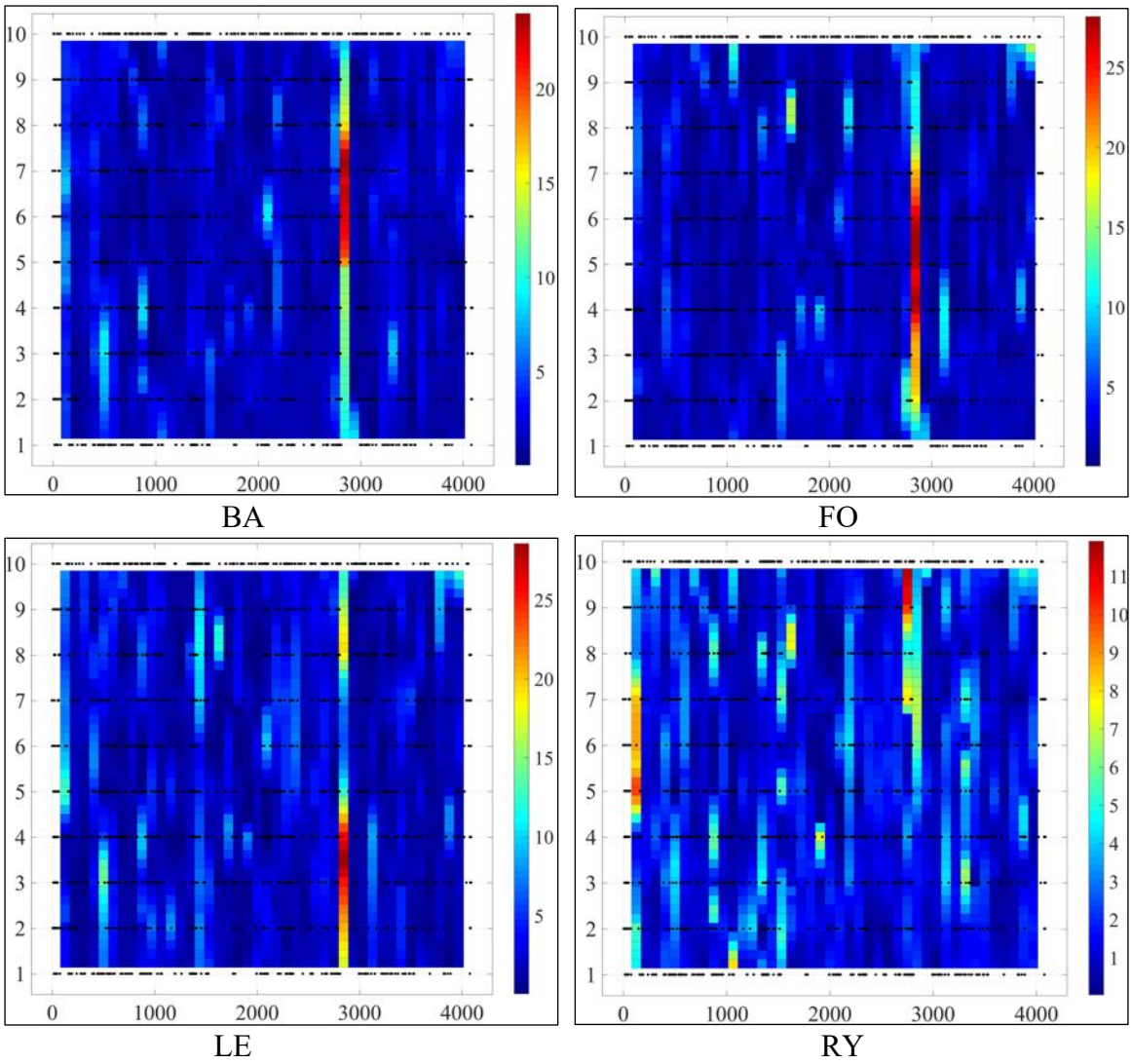



***Figure 4.9:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 4. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

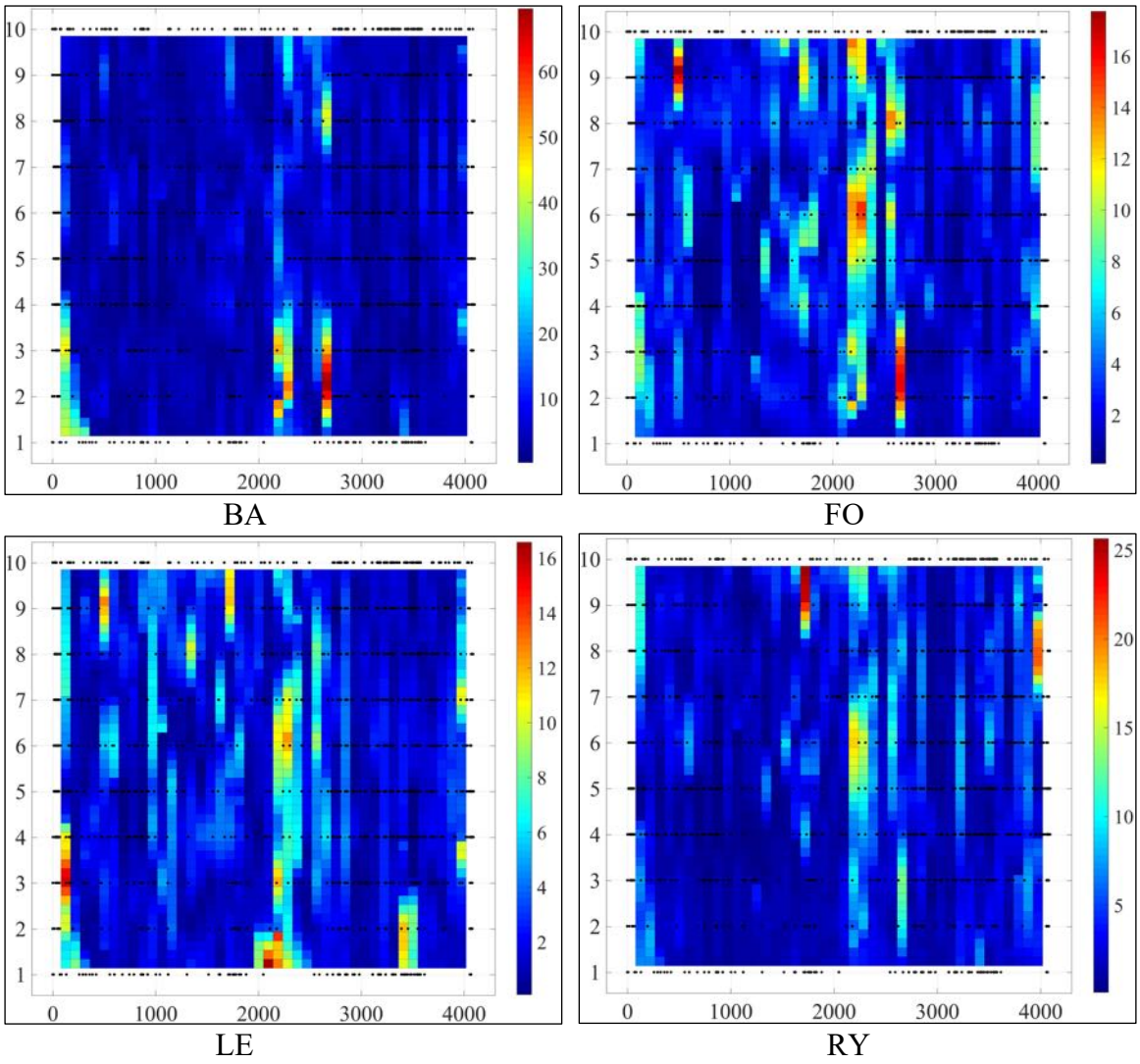



***Figure 4.10:*** *Vectors of the Gabor features from covert speech of one trial, recorded from user 5. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

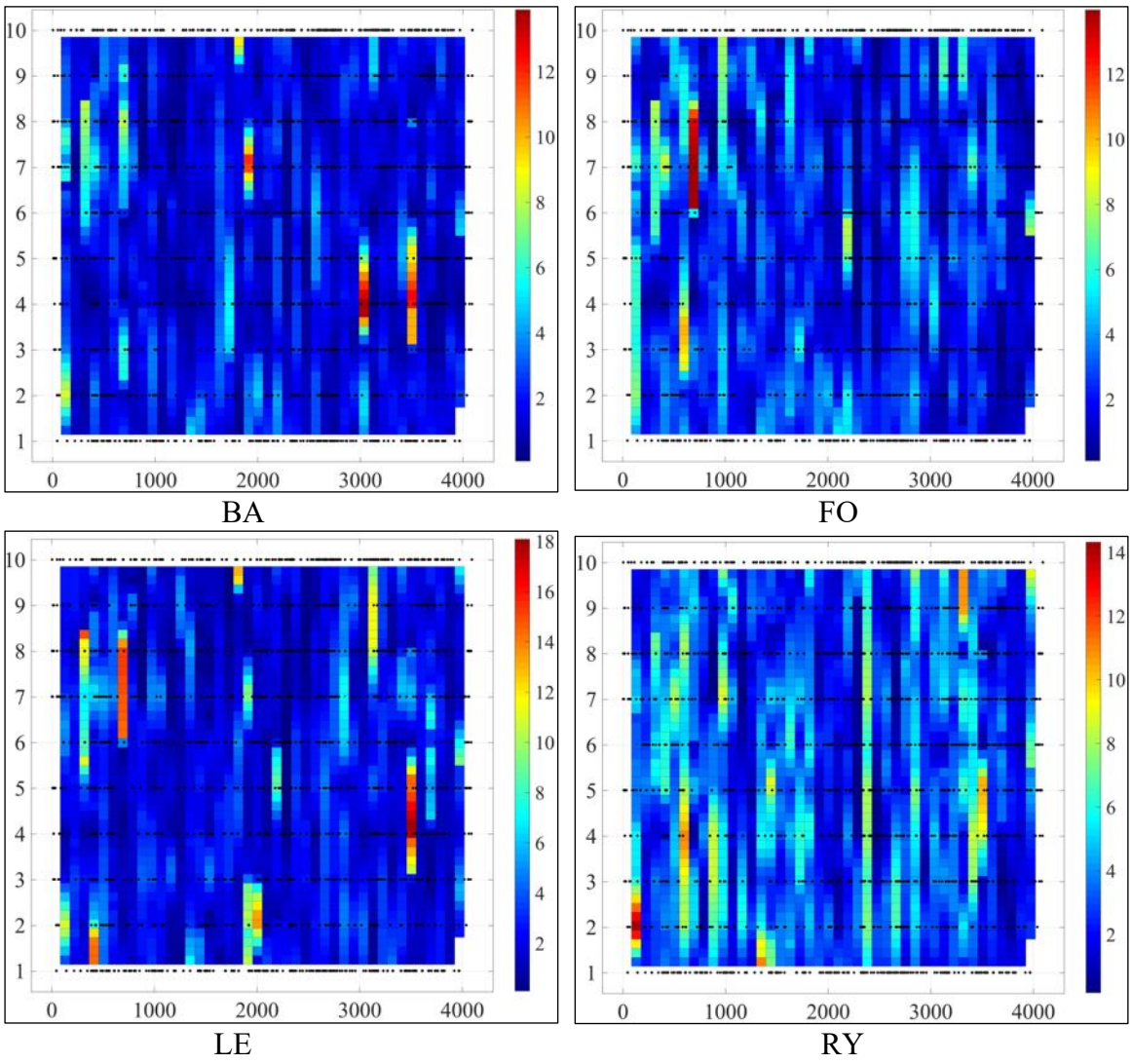



***Figure 4.11:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 6. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

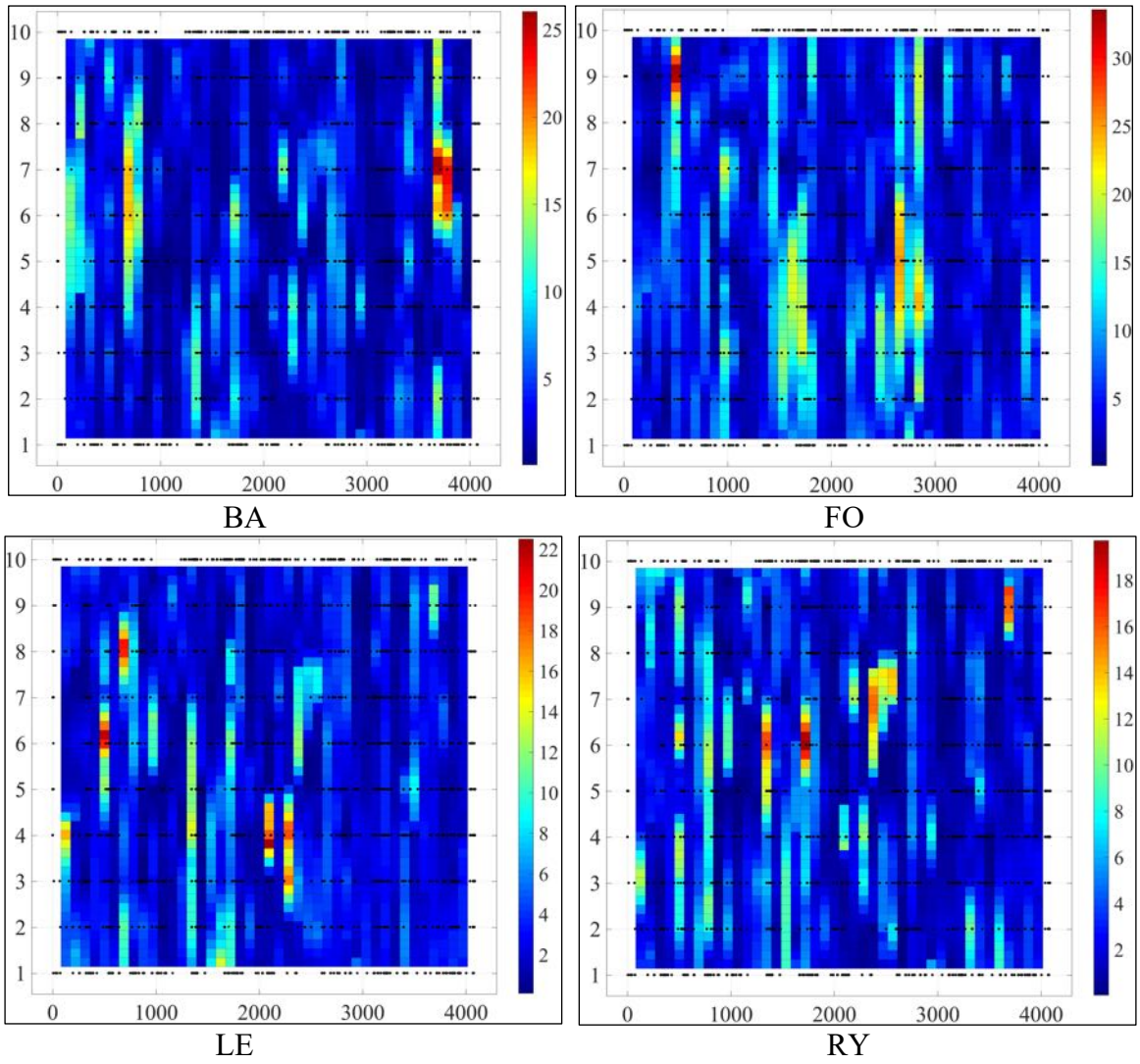



***Figure 4.12:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 7. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

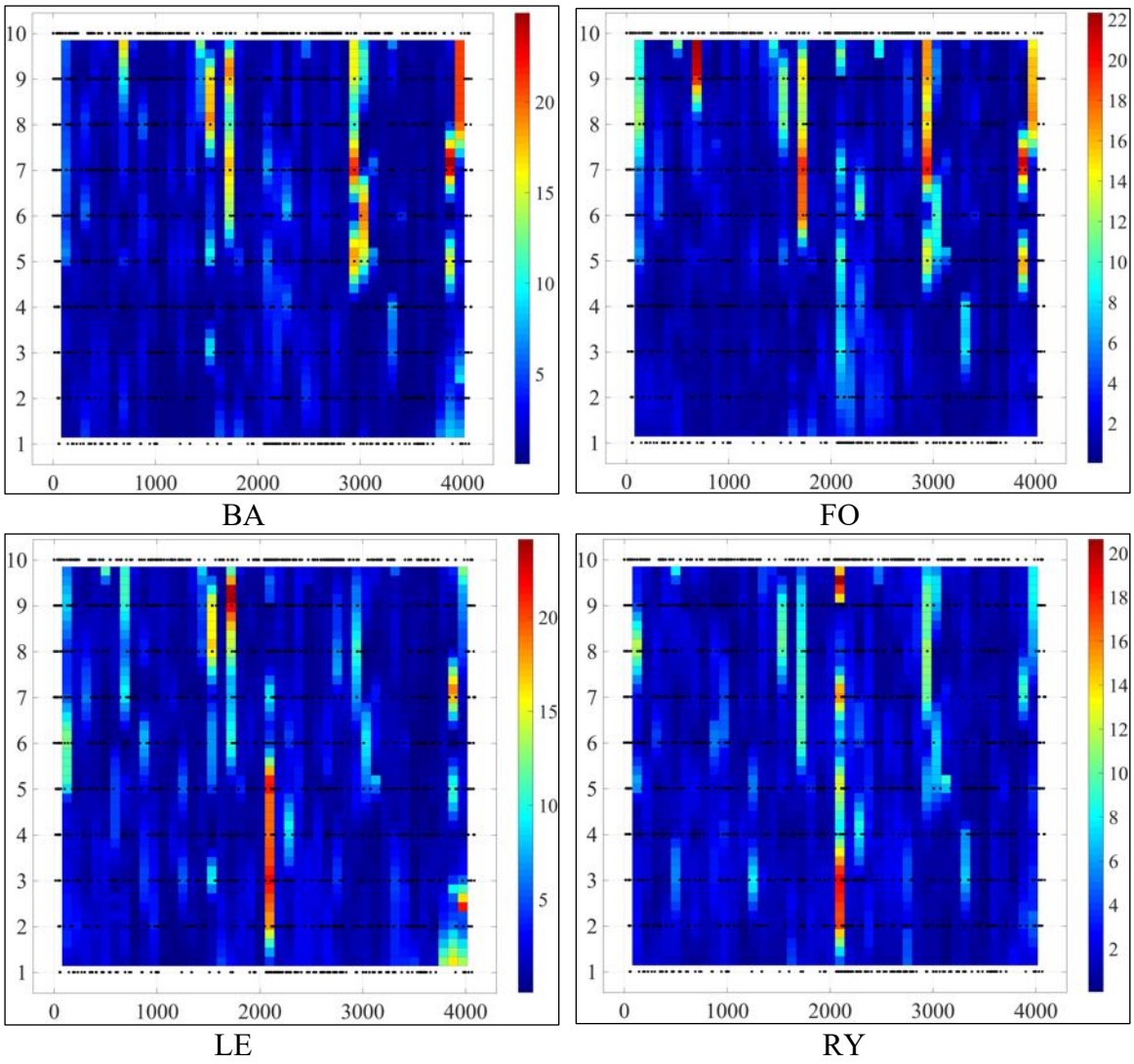



***Figure 4.13:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 8. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

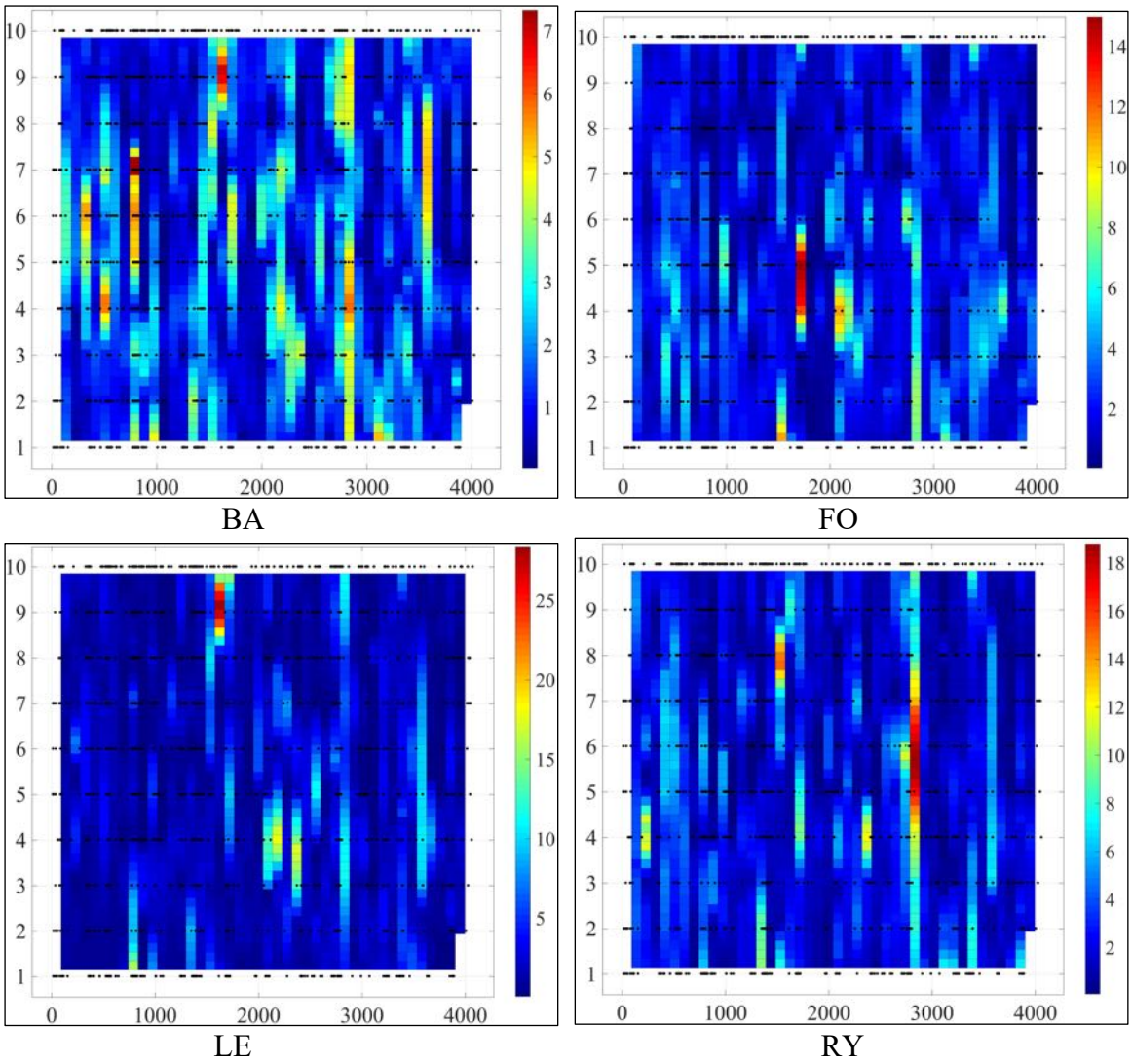



***Figure 4.14:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 9. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

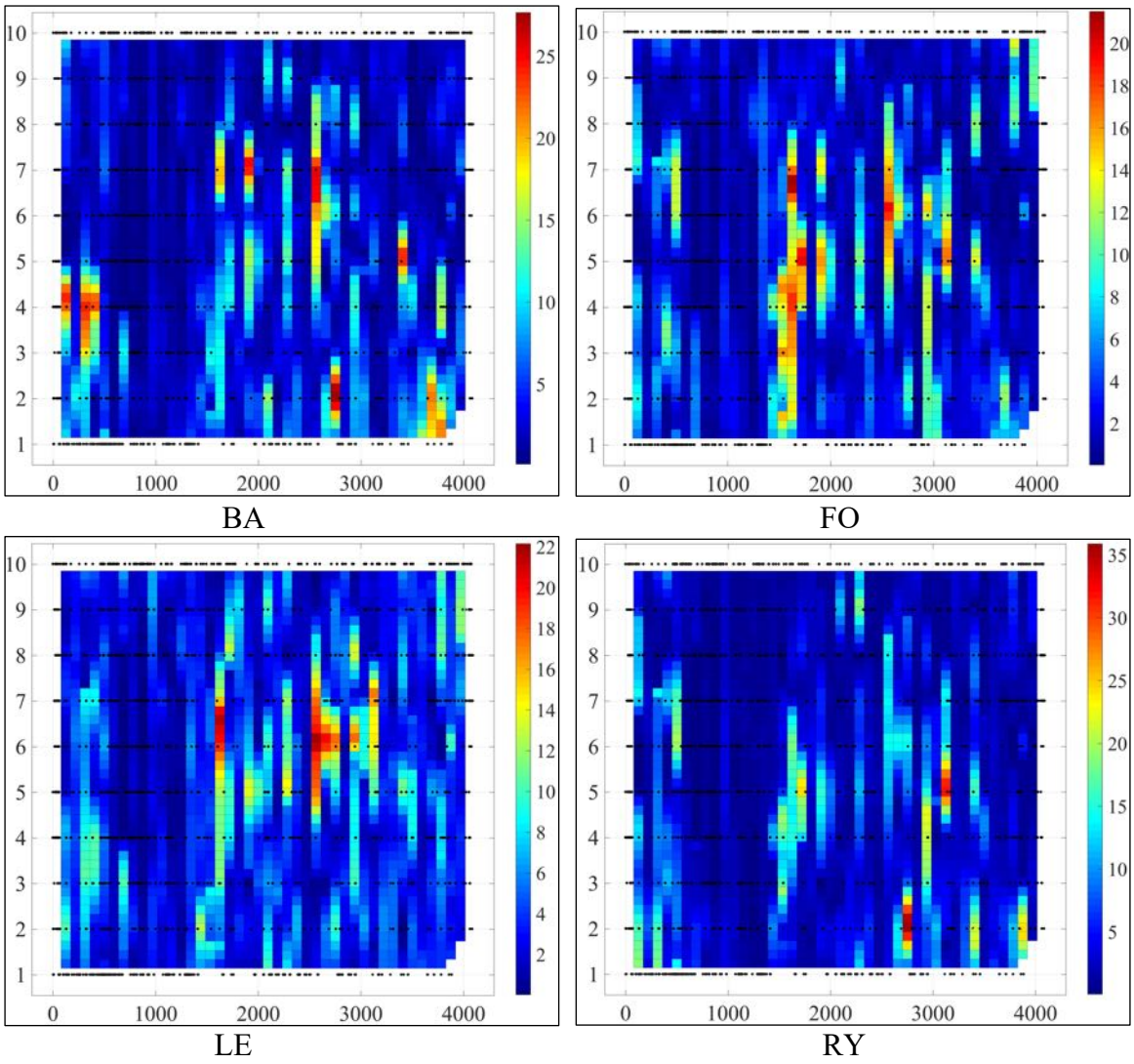



***Figure 4.15:*** *Vectors of the Gabor features from covert speech tasks of one trial, recorded from user 10. The vertical axis has 10 time-steps, and the horizontal axis has 4096 points from 64 channels with 64 frequency bands per channel.*

## *4.4. Time, frequency, and location of Linguistic activity*

Each Gabor feature is linked to a frequency band, time step, and EEG electrode. The outstanding times, frequencies, and locations based on the number of times they have been identified in the most valuable 4K Gabor features, in each of the ten validation folds, for each of the ten users (4e5 in total) are sown in figure 4.16.

Information from the indexes of the 4e5 Gabor features identified in the 312ms trials are cumulatively placed in the 64x10 colour coded time-frequency plot (312ms, 0-128Hz) and used to create a topographical map of the brain based on the location of the EEG electrodes [210, 211].

The most significant regions are the Prefrontal Cortex [160] (stimulus driven executive control), the left Superior Temporal Gyrus [39] (Wernicke's area, phonological code retrieval), the right, and left Inferior Frontal Gyrus [39] (Broca's area, syllabification).

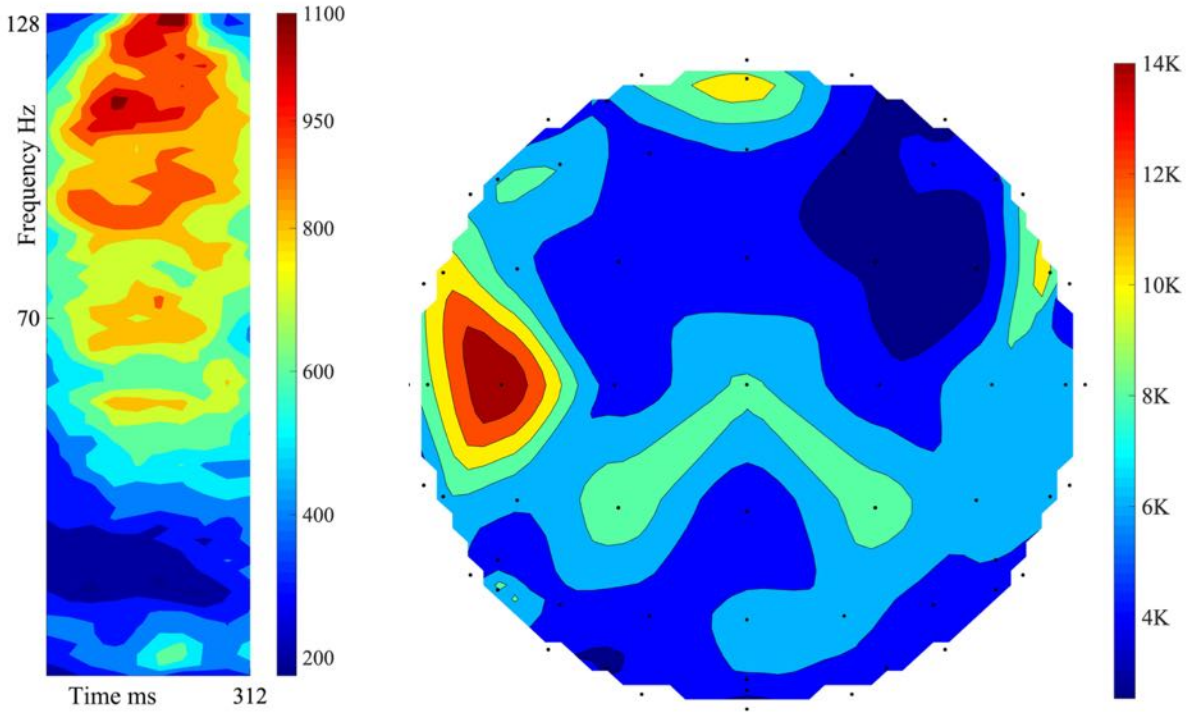



***Figure 4.16:*** *The cumulative colour-coded joint time-frequency representation of 4e5 Gabor features, 312ms trials (Left). The associated topographical plot (Right). Most important regions: Prefrontal Cortex, left STG (Wernicke's area), right, and left IFG (Broca's area).*

Using the same Gabor features as figure 4.16, to demonstrate the sequence of neural activations, topographical plots with 62ms intervals are created and presented in figure 4.17. Each plot only uses the information from Gabor features which lie inside the specified time period.

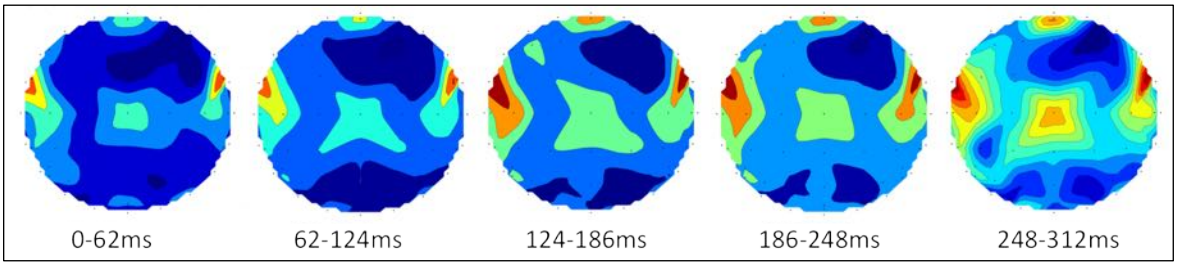



***Figure 4.17:*** *Topographical maps of brain regions generating the most distinctive Gabor features within the indicated 62ms interval. The plot for the 248-312ms interval indicates the early stages of perceptual planning, before activation of the SMA (~500ms) and covert articulation.*

The sequence of activation is as follows [39]:

- [0-62ms] Left, and right Auditory Cortex: response to auditory cue.
- [62-124ms] Prefrontal Cortex [160]: Stimulus-driven executive control, initiating covert speech with auditory cue recognition (100ms). Left Middle Temporal Gyrus: Lemma activation (100-124ms).
- [124-186ms] left Superior Temporal Gyrus: Phonological code retrieval.
- [186-248ms] Left and right Inferior Frontal Gyrus: syllabification.
- [248-312ms] Left inferior, and Superior Parietal Cortex [160]: Goal-driven executive control, by suppressing the Primary Motor Cortex, and activating an internal perceptual planning process [5, 29, 30, 212].

The syllabification stage is completed sooner than estimated, and the 312ms trials contain the very early stages of perceptual planning. However, the covert articulation stage, which occurs after the activation of the Supplementary Motor Area [39, 213], is excluded from shortened trials as intended. In the 312ms trials, the spatial, temporal, and spectral properties of the 4e5 most valuable Gabor features identified from 10 participants, correspond to the automatic linguistic processing stages of word production prior to articulation, and are supported by a substantial body of evidence [23-25, 27, 39-43, 46, 47, 187, 212].

## *4.5. Discussion*

In chapter three, trials were recorded in a mixed randomised sequence [214]. In this chapter, trials were recorded in 4 blocks (30 trials of one class per block) to reduce recording time (7-8 mins) and prevent user fatigue, as recording 120 trials in a single run, required 25-30 minutes.

To demonstrate there are no drifts in the EEG signal (i.e. change of an electrode's impedance) , the distribution properties (mean, STD, rang, ...) of the raw EEG recordings for each user should be virtually identical in all four blocks. This is indeed the case. Figure 4.18 presents the distribution properties of the recorded blocks from user 1.

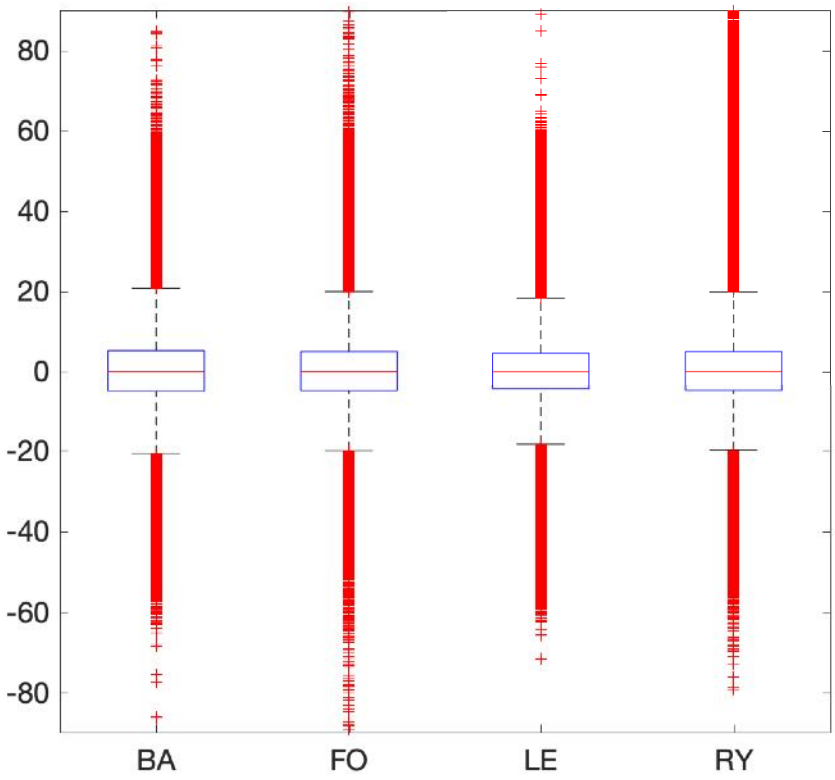



***Figure 4.18:*** *The distribution properties of raw EEG recordings in each block for user 1. In all blocks, the mean is 0, STD is 10, the 25% and 75% quartiles are -20 and 20 respectively, and range is near 180. They all have Gaussian distribution. With classification accuracy of 96.7%, no signs of signal drifting exist, suggesting that recording in blocks has little, if any effect on classification accuracy for this data.*

Another issue can arise from the use of 10-fold cross validation, is dependencies between trials recorded in blocks. An alternative to 10-fold cross validation is HV-cross-validation, which uses large validation sets to sufficiently separate the training trials from the testing trials recorded in blocks [215]. Figure 4.19 illustrates the application of HV-cross-validation for the 30 trials recorded in blocks in this chapter. For each fold, 21 trials are used for training, 3 trials are used for testing, and six trials are used as validation trials to separate the training trials from the testing trials.

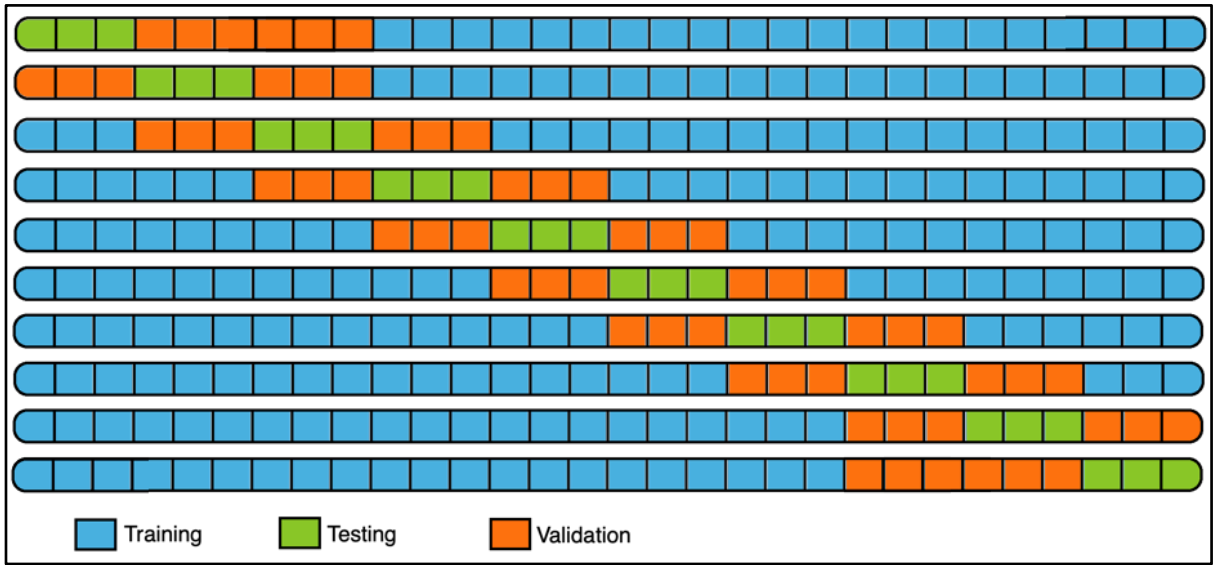



***Figure 4.19:*** *Alternative to ten-fold cross-validation known as HV-cross-validation. In this method, an additional validation buffer zone (6 trials) is excluded from the training set (21 trials) to remove trials that are closest to the testing set (3 trials) and reduce the possible effects of temporal correlation between training and testing sets. All 30 trials are used at least once for testing.*

To reject the hypothesis that the results were achieved by chance, the analysis was also performed for randomised labels, and for pre-cue idle state, for all participants. Table 4.2 presents the results of HV-cross validation, randomised labels, and pre-cue rest state, compared to the original results in the first column.

| | *10-fold cross validation* | *HV cross validation* | *Randomised labels* | *312ms pre-cue (idle state)* |
|---|---|---|---|---|
| *User 1* | *96.7±7* | *84.1±15.4* | *25.8±13.2* | *28.3±8* |
| *User 2* | *94.1±8.8* | *92.5±10.7* | *31.6±14.5* | *19.1±18.8* |
| *User 3* | *100* | *91.6±9.6* | *24.1±13.2* | *25.8±15.4* |
| *User 4* | *97.5±5.6* | *98.3±3.5* | *22.5±9.6* | *24.1±7.2* |
| *User 5* | *92.5±10* | *81.6±16.5* | *27.5±13.6* | *24.2±9.2* |
| *User 6* | *89.1±9.6* | *65.8±19* | *25±10.3* | *20±9.8* |
| *User 7* | *91.6±11.1* | *94.1±8.8* | *18.3±11.6* | *20.8±8* |
| *User 8* | *89.1±8.8* | *80±13.1* | *23.3±11.6* | *21.7±13.7* |
| *User 9* | *100* | *97.5±7.9* | *16.6±11.1* | *25±13* |
| *User 10* | *94.1±7.9* | *89.1±13* | *25±9.6* | *26.6±12.2* |
| *Average* | *94.5±4* | *87.4±11.7* | *24±11.8* | *23.5±11.5* |

***Table 4.2:*** *The results of using HV-cross validation, randomised labels, and pre-cue rest state, compare to the original reported results in the first column.*

As seen in table 4.2, it is clear that the original ten-fold cross validation was upwardly biased. The HV-cross-validation scheme, which uses validation folds to separate the testing trials from training trials, is more realistic. The upward bias is a result of using block recordings. By comparing the performance of the online system in the next chapter and comparing that to the estimated performance using HV-cross validation in table 4.2, the exact amount of upward bias can be determined.

From table 4.2, the performance of the system using randomised labels, and pre-cue rest date is close to chance level. The hypothesis that the reported results were achieved by chance is rejected. In addition to the pre-cue rest data, trials from the resting state after covert speech tasks are used for classification and presented in table 4.3. The performance for these trials is also near chance level, which strengthen the above conclusion.

| | *1 to 1.312 sec* | *2 to 2.312 sec* | *3 to 3.312 sec* |
|---|---|---|---|
| *User 1* | *27.5±16.2* | *27.5±10.4* | *35±5.5* |
| *User 2* | *20±6.8* | *12.5±12.5* | *22.5±18.5* |
| *User 3* | *17.5±11.2* | *22.5±16.2* | *10±5.6* |
| *User 4* | *20±18.9* | *35±10.4* | *22.5±20.5* |
| *User 5* | *32.5±14.2* | *22.5±5.6* | *30±11.2* |
| *User 6* | *22.5±10.4* | *10±10.5* | *25±8.8* |
| *User 7* | *17.6±14.2* | *15±13.6* | *10±5.5* |
| *User 8* | *20±14.2* | *32.5±14.3* | *25±8.8* |
| *User 9* | *35±13.6* | *15±5.5* | *32±14.2* |
| *User 10* | *22.5±13.6* | *23±10.4* | *25±8.8* |
| *Average* | *23.5±13.3* | *21.5±10.9* | *23.7±10.7* |

***Table 4.3:*** *Time-coarse of classification accuracy after task trial, during the rest state before the presentation of next task. The performances are near chance level.*

Table 4.4 presents the classification accuracy for fewer valuable Gabor features. As seen, the performance increases steadily as the number of features grow, and no evidence of overfitting is observed.

| | *30 features* | *100 features* | *300 features* | *1K features* | *2K features* | *3K features* | *4K features* |
|---|---|---|---|---|---|---|---|
| *User 1* | *68.3±10.9* | *35±16.5* | *84.1±11.4* | *95.8±5.8* | *98.3±5.2* | *98.3±5.2* | *96.7±7* |
| *User 2* | *55±18* | *40.8±14.9* | *83.3±15.2* | *90.8±12.6* | *93.3±10.24* | *94.1±10.4* | *94.1±8.8* |
| *User 3* | *65.8±14.9* | *49.1±15.9* | *86.6±10.5* | *93.3±8.6* | *90.8±8.2* | *91.6±9.6* | *100* |
| *User 4* | *65±17.9* | *54.1±14.8* | *94.1±7.9* | *96.6±8* | *98.3±3.5* | *98.3±3.51* | *97.5±5.6* |
| *User 5* | *51.6±20.3* | *45.8±11.9* | *75.8±14.9* | *80.8±11.8* | *89.1±11.1* | *91.7±9.6* | *92.5±10* |
| *User 6* | *30.8±10.4* | *37.5±10.5* | *64.1±11.8* | *75.8±6.1* | *82.5±11.4* | *85.8±5.6* | *89.1±9.6* |
| *User 7* | *51.6±14.5* | *41.6±15.2* | *85±12.2* | *90±10.2* | *94.1±8.8* | *94.2±8.8* | *91.6±11.1* |
| *User 8* | *56.6±18.3* | *40±14.5* | *78.3±15.3* | *81.6±13.4* | *85.8±11.1* | *86.6±9.8* | *89.1±8.8* |
| *User 9* | *70.8±13.7* | *40±16.6* | *96.6±8* | *97.5±7.9* | *97.5±7.9* | *97.5±7.9* | *100* |
| *User 10* | *50.8±11.4* | *30.8±20* | *80.8±13.6* | *89.1±10.4* | *94.1±8.8* | *94.1±8.8* | *94.1±7.9* |
| *Average* | *41.5±15* | *56.6±15* | *82.9±12* | *89.1±9.5* | *92.4±8.6* | *93.2±7.9* | *94.5±4* |

***Table 4.4:*** *Classification accuracy using fewer valuable Gabor features*

Finally, to completely reject the Null Hypothesis, for each user, within each validation fold, 100 randomised lable tests are performed. the average of these 100 randomised tests are compare with the correct lable test. These are shown in table 4.5. The grand average spectrogram for each user for each class in the central channel Cz are presented in figures 4.20 to 4.29. There are clear differences in the plots for each class for all users. Although feature power is not used for classification and the lower frequency bands are much more visible in the plots, there are still clear differences in the plots for each class. The rejection of null hypothesis and the grand average plots are completely aligned with the reported resuts and provide additional validation .

| | | CV1 | CV2 | CV3 | CV4 | CV5 | CV6 | CV7 | CV8 | CV9 | CV10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U1 | Correct | 58 | 83.3 | 83.3 | 91.7 | 100 | 91.7 | 100 | 91.7 | 83.3 | 66.7 | P=1.73e-4 |
| | Random | 24.65 | 26.4 | 23.2 | 25.75 | 22.65 | 27.4 | 22.15 | 23.85 | 27.2 | 24.9 | h=1 |
| U2 | Correct | 91.7 | 91.7 | 91.7 | 100 | 100 | 100 | 100 | 83.3 | 100 | 66.7 | P=1.59e-4 |
| | Random | 27.25 | 26.8 | 25.15 | 31.2 | 24.65 | 30.75 | 24.35 | 28.9 | 27.35 | 29.2 | h=1 |
| U3 | Correct | 83.3 | 100 | 100 | 100 | 100 | 100 | 83.3 | 83.3 | 75 | 91.7 | P=1.59e-4 |
| | Random | 25.6 | 22.45 | 29.85 | 27.15 | 23.35 | 25.25 | 21.6 | 24.65 | 24.2 | 23.75 | h=1 |
| U4 | Correct | 100 | 100 | 91.7 | 100 | 100 | 100 | 100 | 100 | 100 | 91.7 | P=1.09e-4 |
| | Random | 20.95 | 19.8 | 24.45 | 23.25 | 18.6 | 22.75 | 20.65 | 23.45 | 22.85 | 23.15 | h=1 |
| U5 | Correct | 58.3 | 100 | 100 | 83.3 | 100 | 91.7 | 75 | 83.3 | 66.7 | 58.3 | P=1.77e-4 |
| | Random | 28.15 | 25.6 | 22.35 | 31.65 | 24.55 | 28.4 | 28.35 | 24.2 | 26.95 | 27.65 | h=1 |
| U6 | Correct | 58.3 | 50 | 91.6 | 58.3 | 100 | 58.3 | 83.3 | 58.3 | 58.3 | 41.7 | P=1.63e-4 |
| | Random | 23.1 | 25.75 | 26.45 | 20.55 | 27.6 | 28.35 | 25.3 | 27.85 | 24.3 | 21.25 | h=1 |
| U7 | Correct | 100 | 100 | 100 | 100 | 100 | 91.7 | 100 | 91.7 | 83.3 | 75 | P=1.48e-4 |
| | Random | 18.3 | 20.65 | 24.2 | 17.65 | 20.4 | 16.55 | 19.4 | 17.35 | 19.15 | 20.85 | h=1 |
| U8 | Correct | 91.7 | 83.3 | 83.3 | 91.7 | 75 | 75 | 91.7 | 91.7 | 58.3 | 58.3 | P=1.69e-4 |
| | Random | 24.85 | 29.3 | 23.25 | 19.75 | 22.6 | 24.55 | 26.75 | 21.4 | 25.85 | 22.4 | h=1 |
| U9 | Correct | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 75 | P=8.68e-5 |
| | Random | 19.65 | 20.9 | 17.35 | 19.65 | 22.7 | 21.45 | 18.3 | 24.85 | 23.65 | 18.7 | h=1 |
| U10 | Correct | 75 | 91.7 | 100 | 100 | 100 | 100 | 100 | 83.3 | 75 | 66.7 | P=1.62e-4 |
| | Random | 27.4 | 22.65 | 20.7 | 28.15 | 24.4 | 25.25 | 25.1 | 24.65 | 22.85 | 25.15 | h=1 |

***Table 4.5:*** *Rejecting Null Hypothesis: For each user, within each of the 10 cross-validation folds, the performance is tested with randomised labels 100 times. The average of these 100 tests with randomised labels is shown under the performance with correct labels for each cross-validation fold. Wilcoxon rank-sum tests are performed. The value of h=1 rejects the null hypothesis suggesting the results are indeed statistically significant and have not been achieved by chance.*

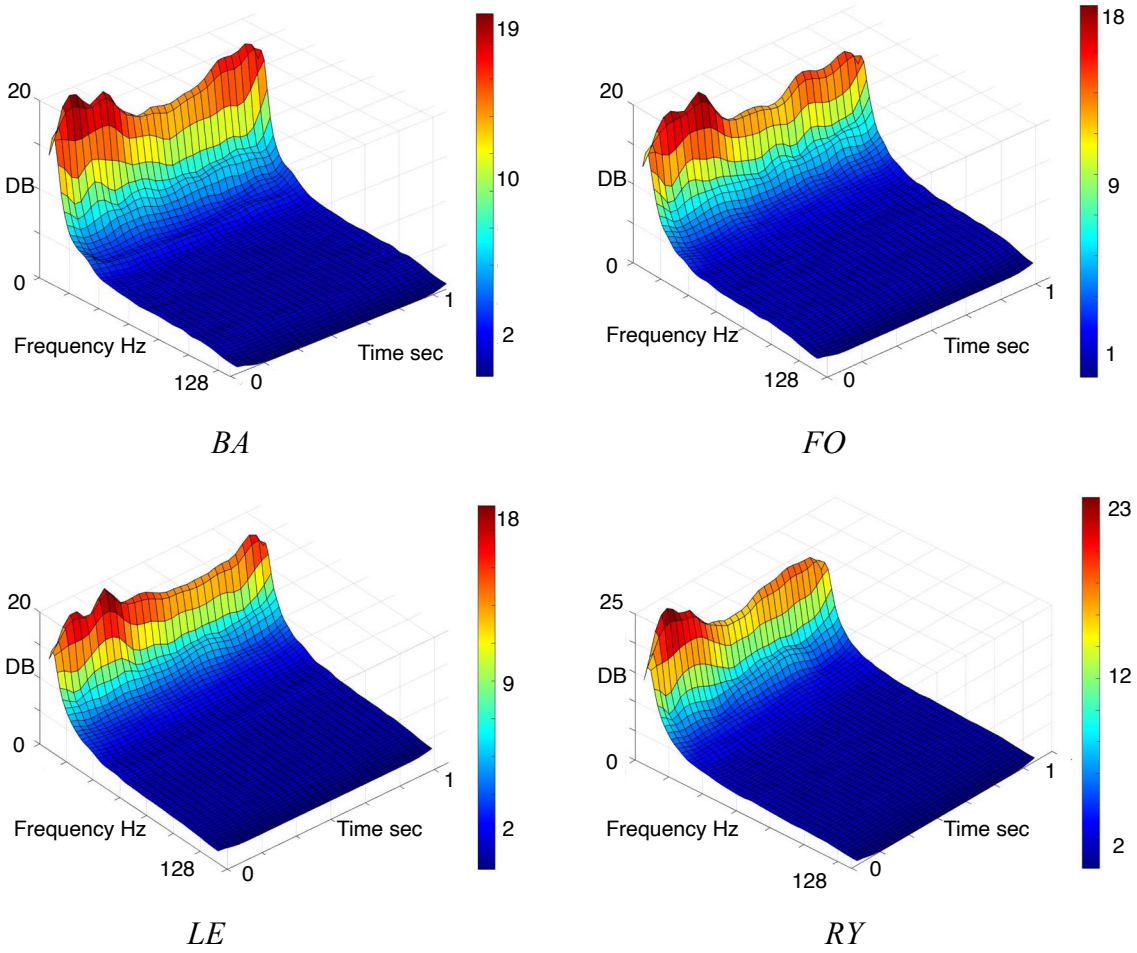



***Figure 4.20:*** *User 1: Grand average spectrogram of each class from all 30 trials and all channels.*

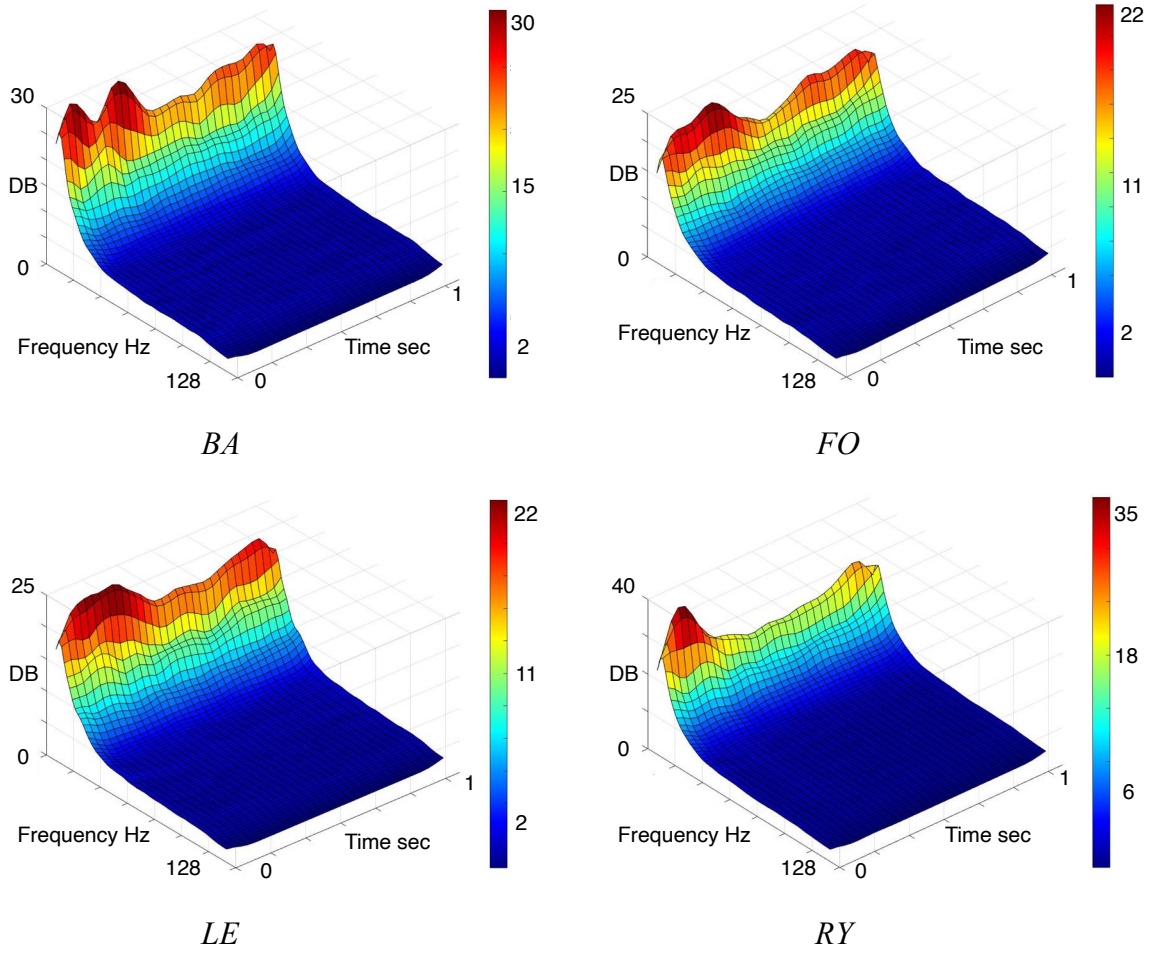



***Figure 4.21:*** *User 2: Grand average spectrogram of each class from all 30 trials and all channels.*

| | N1 | N2 | N3 | N4 |
|---|---|---|---|---|
| BA | (3,1,29.34) | (9,1,28.86) | (30,1,24.93) | (32,3,22.59) |
| FO | (3,1,19.5) | (9,1,21.85) | (30,1,18.67) | (32,20,17.28) |
| LE | (3,1,20.96) | (9,1,22.69) | (30,1,20.15) | (32,20,16.39) |
| RY | (3,1,32.08) | (9,1,22.37) | (30,1,24.98) | (32,3,20.07) |

| P (BA,FO) | P (BA,LE) | P (BA,RY) | P (LE,FO) | P (LE,RY) | P(RY,FO) | P TOTAL |
|---|---|---|---|---|---|---|
| 0.0286 | 0.029 | 0.057 | 0.068 | 0.067 | 0.0342 | 0.0473 |

***Table 4.6:*** *The total P-value is calculated as the average of binary P-values. In addition to visible class differences in the Spectrogram plots, the small total P-value also supports class separability.*

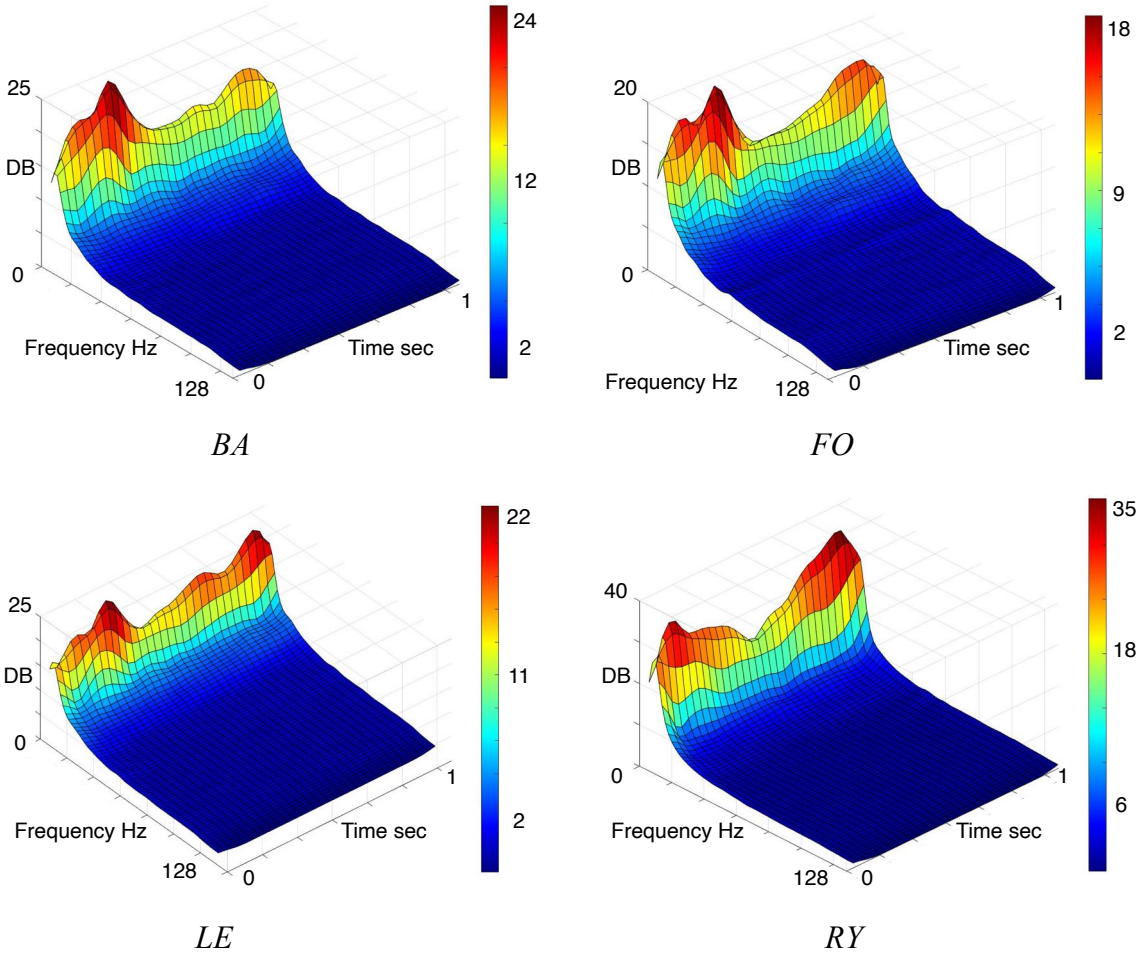



***Figure 4.22:*** *User 3: Grand average spectrogram of each class from all 30 trials and all channels.*

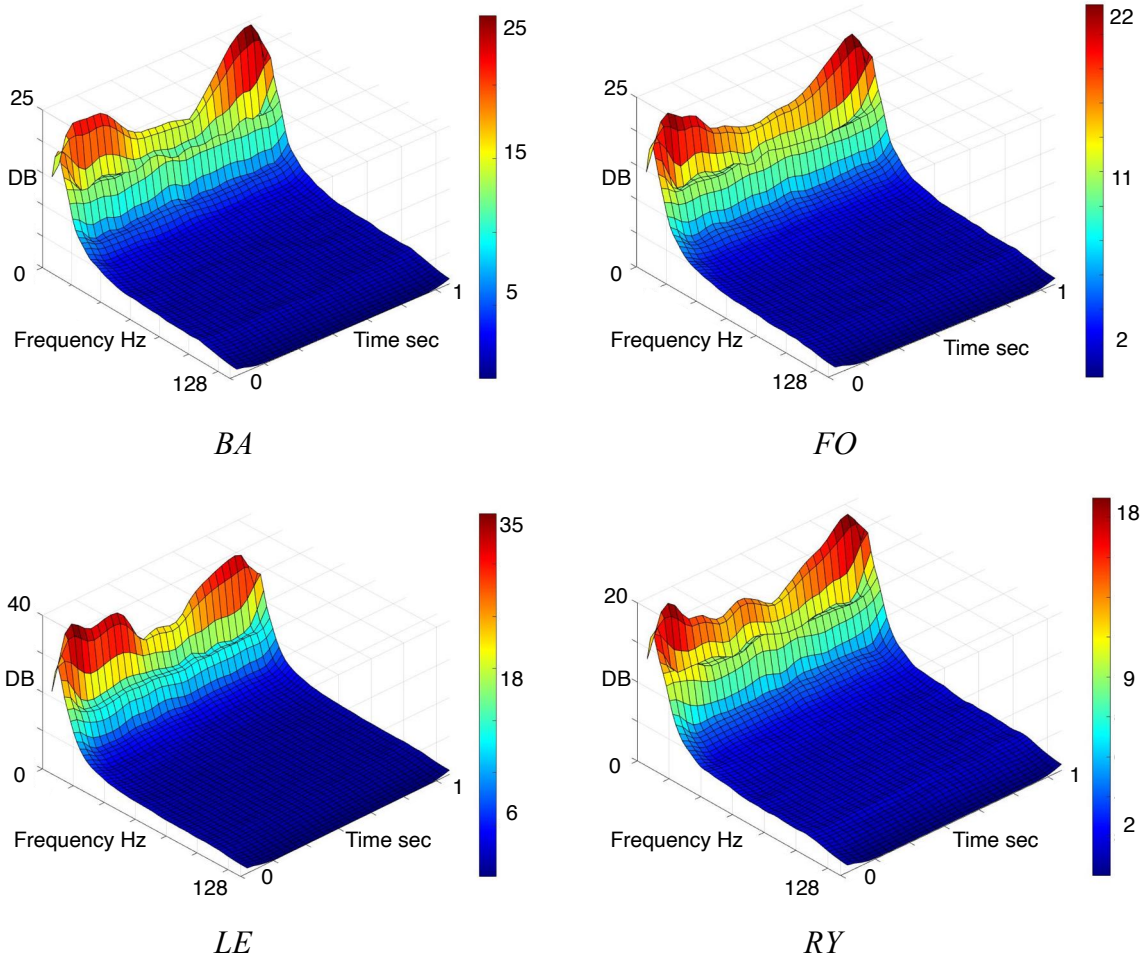



***Figure 4.23:*** *User 4: Grand average spectrogram of each class from all 30 trials and all channels.*

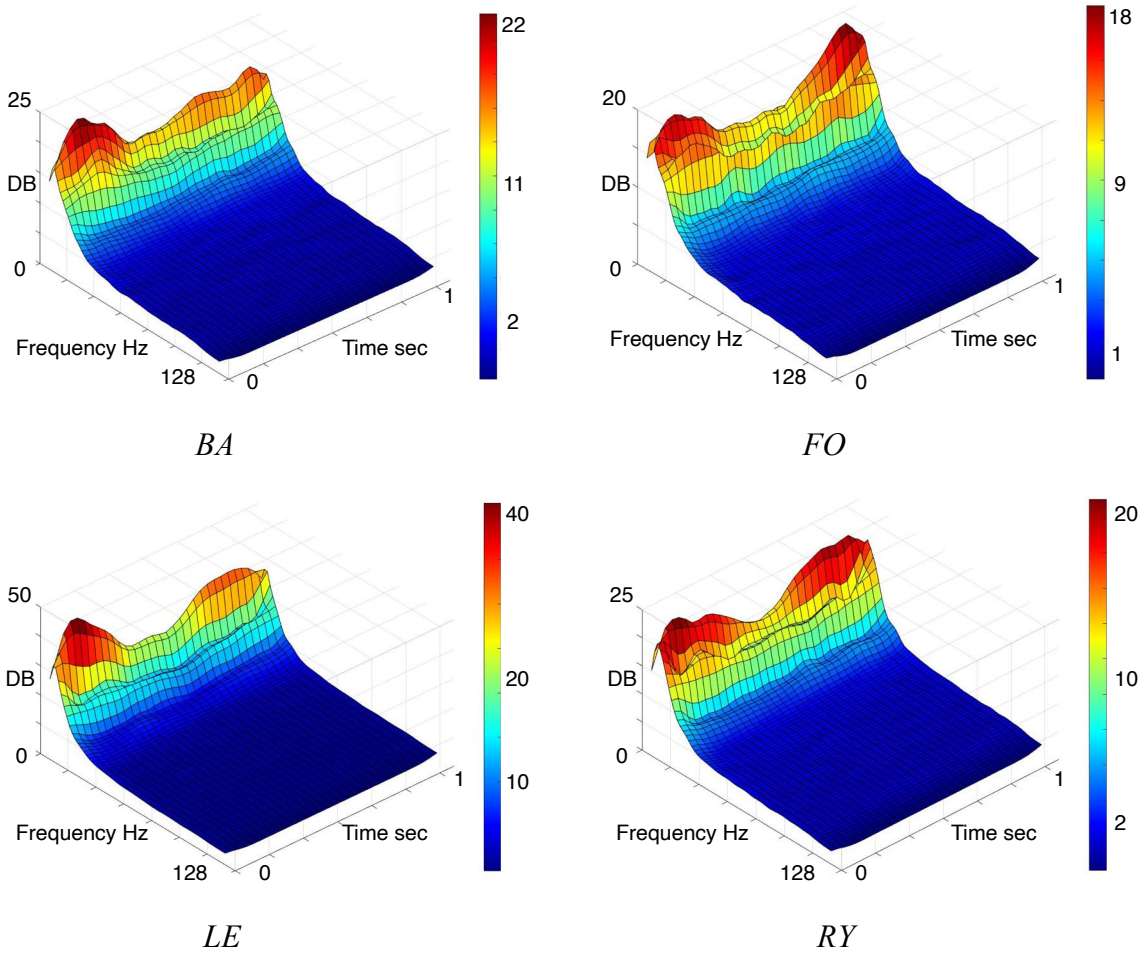



***Figure 4.24:*** *User 5: Grand average spectrogram of each class from all 30 trials and all channels.*

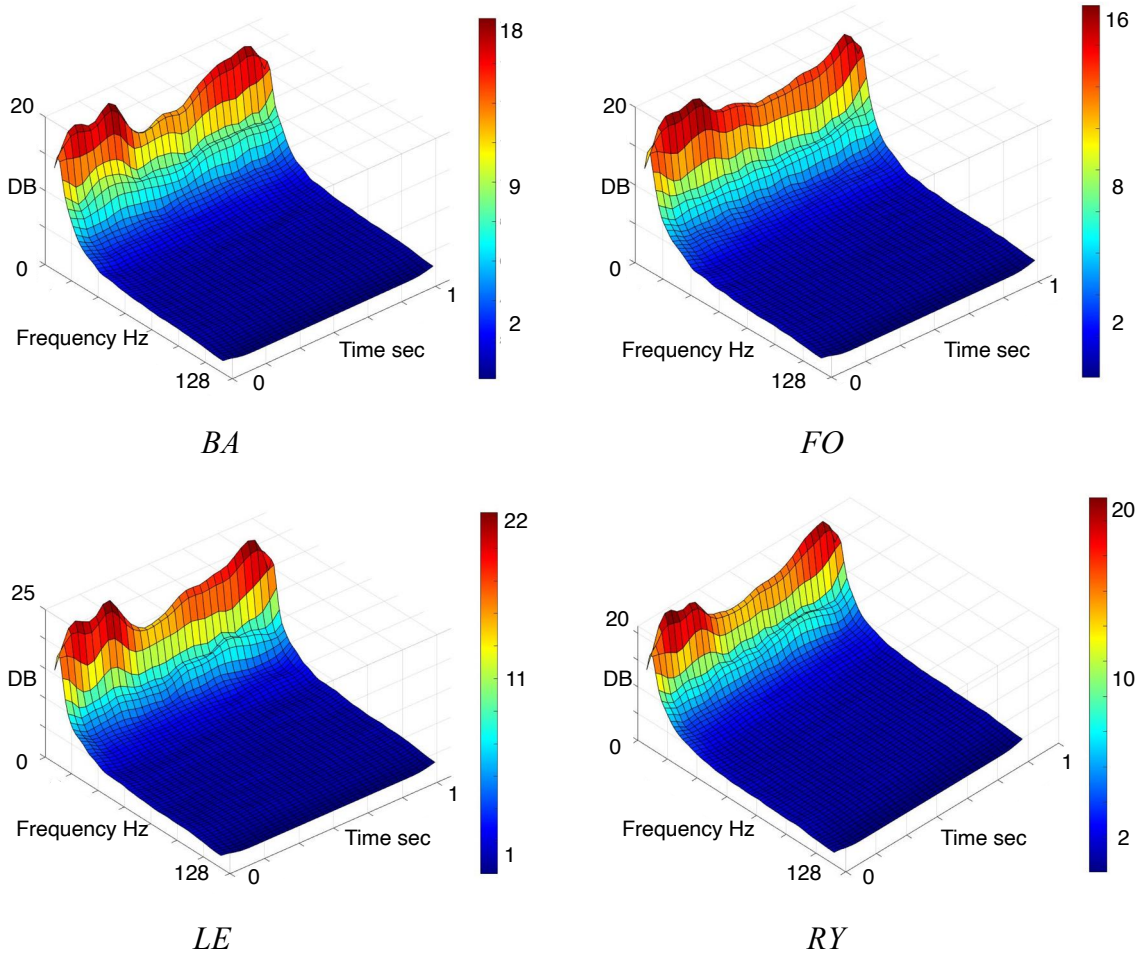



***Figure 4.25:*** *User 6: Grand average spectrogram of each class from all 30 trials and all channels.*

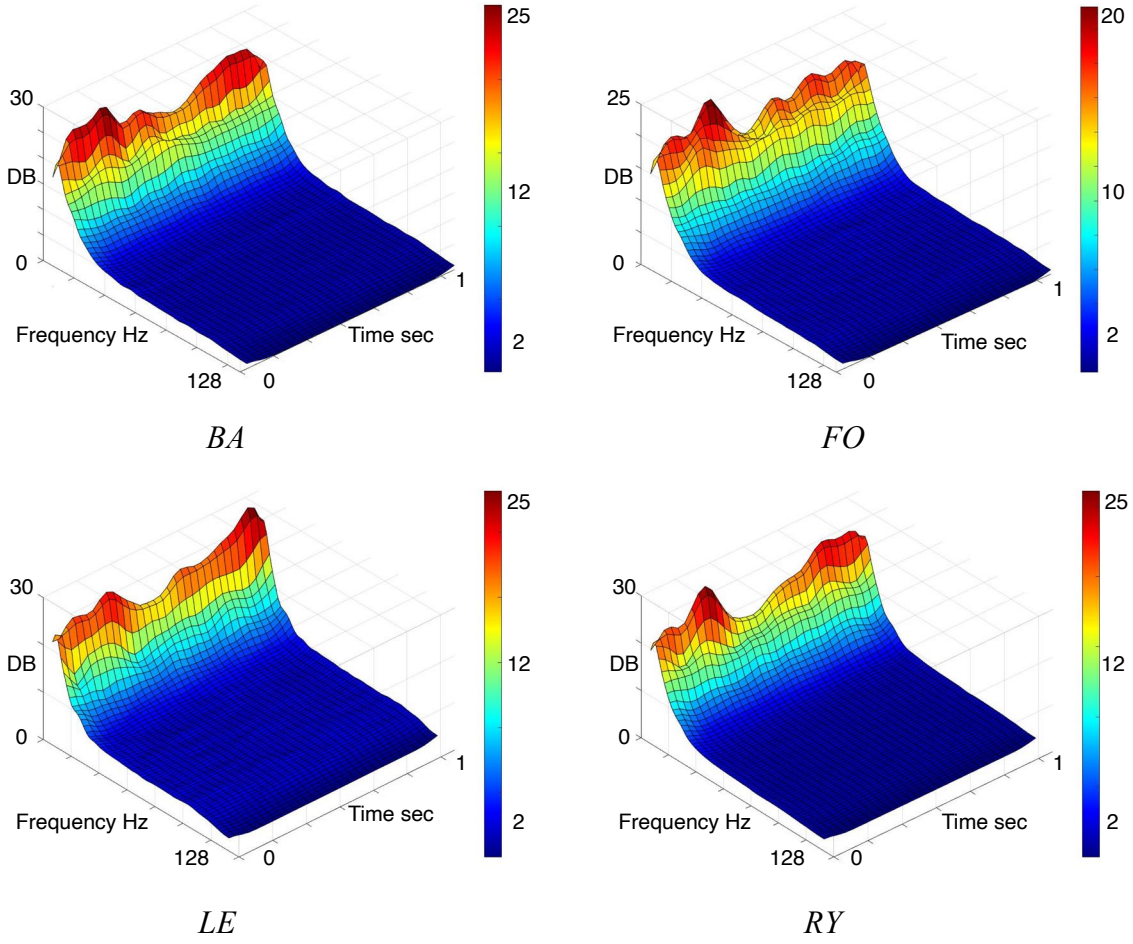



***Figure 4.26:*** *User 7: Grand average spectrogram of each class from all 30 trials and all channels.*

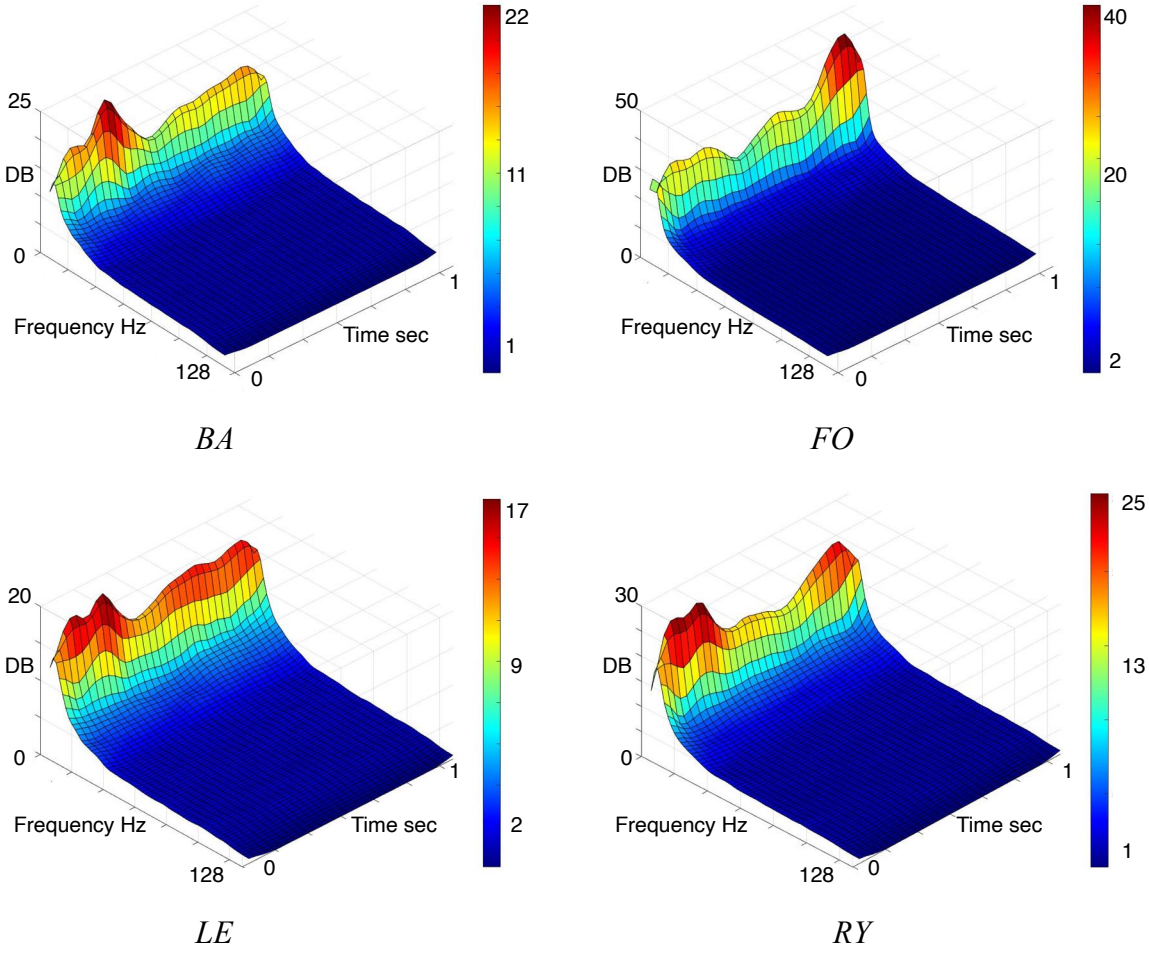



***Figure 4.27:*** *User 8: Grand average spectrogram of each class from all 30 trials and all channels.*

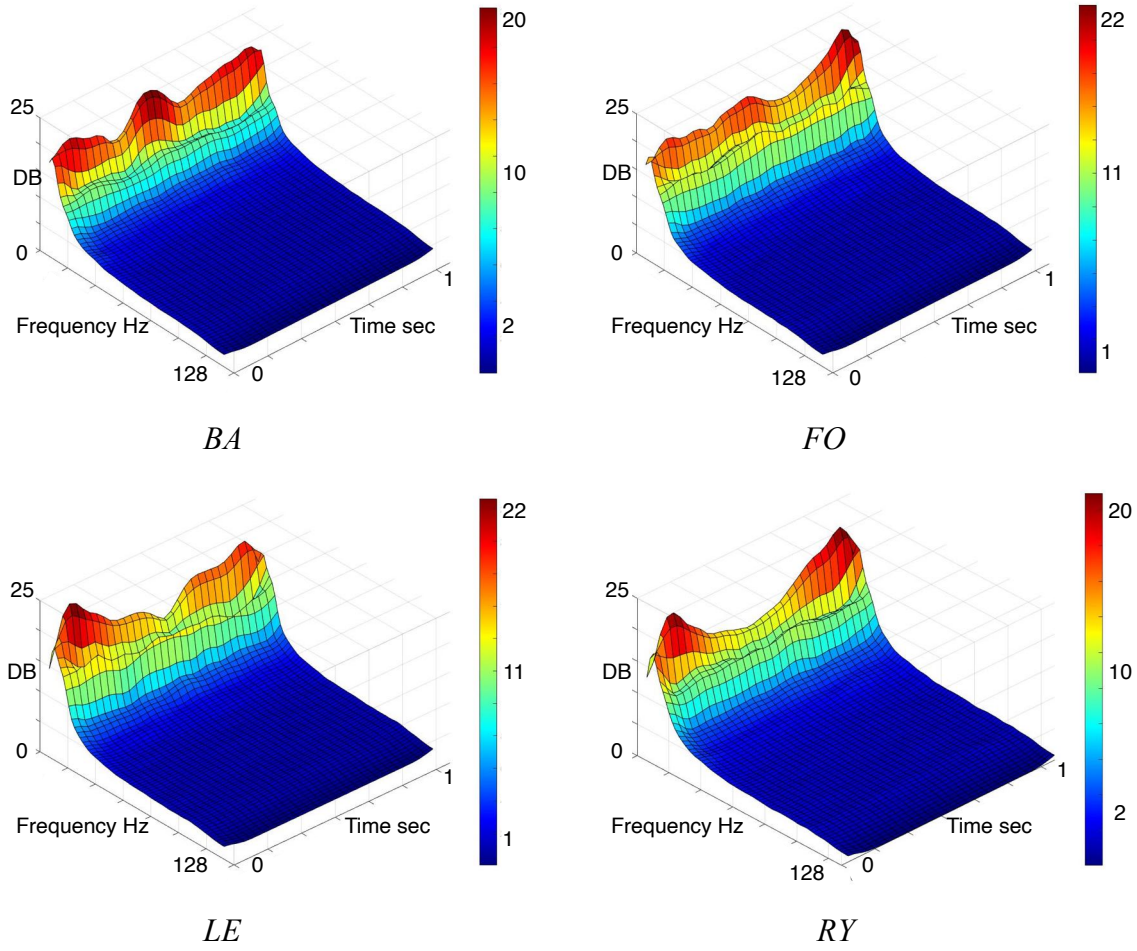



***Figure 4.28:*** *User 9: Grand average spectrogram of each class from all 30 trials and all channels.*

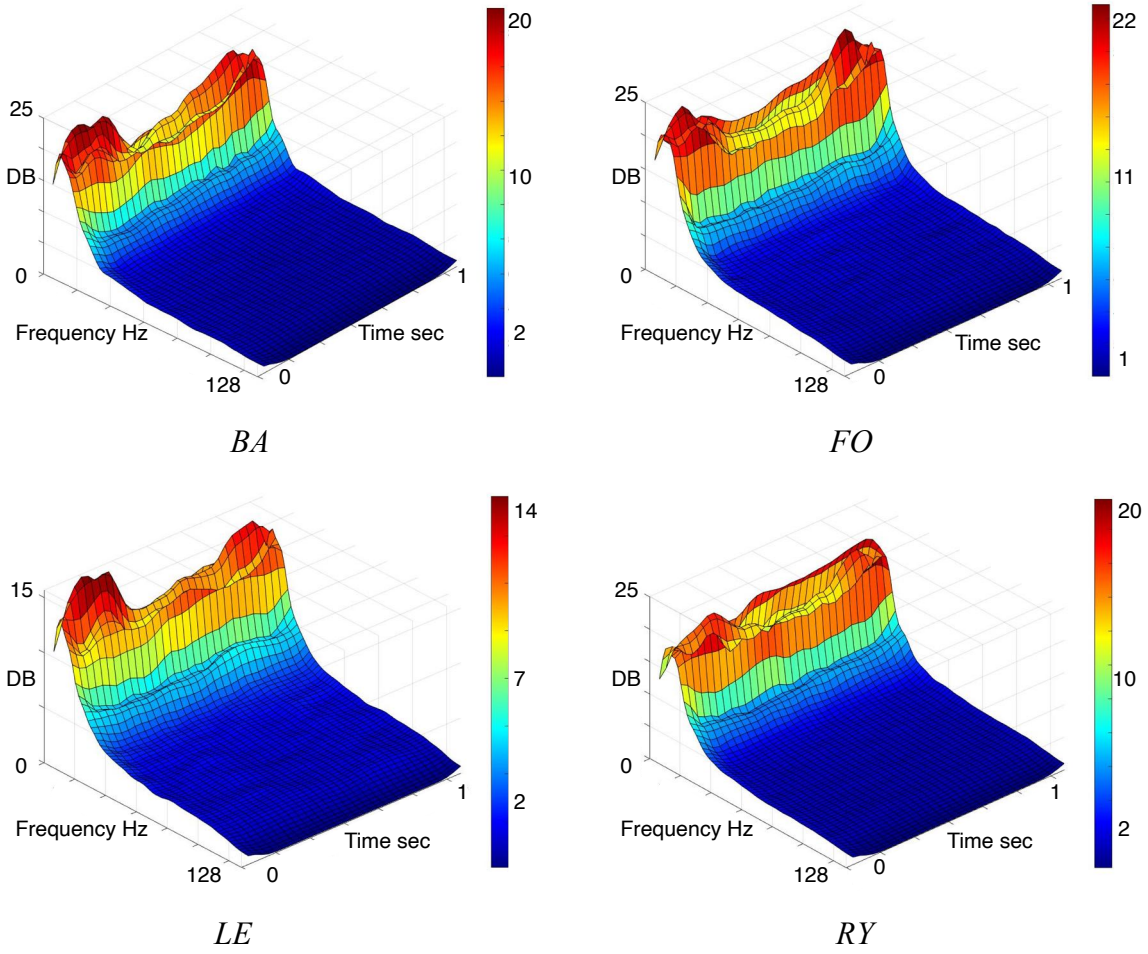



***Figure 4.29:*** *User 10: Grand average spectrogram of each class from all 30 trials and all channels.*

## *4.6. Best settings for online system*

Phonetic linguistic processing prior to articulation, elicits a unique and word-specific pattern of high-Gamma activity [40, 216], which does not change over time [42, 43] and is not affected by frequency [44] or priming [45]. Phonetic codes are set up and consolidated with the acquisition of language during childhood, and remain unchanged throughout a person's life [45]. Phonetic codes are stored in the long term memory, and are processed automatically by the brain requiring no conscious effort from the user during trials, with immunity from any influence or modification [44, 45, 188, 216]. The experimental protocol and analysis pipeline for 312ms trials presented in this chapter is used as a framework to create an online EEG-based 4-class linguistic BCI in the next chapter.

# CHAPTER 5:
## ONLINE APPLICATION OF BCI

## *5.1. Design of the Online Linguistic BCI*

The online Linguistic BCI presented in this chapter, is in fact based on the experimental protocol and analysis pipeline for 312ms trials, presented in chapter four. This experiment was conducted with the help of six neurologically healthy volunteers in the age group of 21-33. All volunteers signed a consent form based on the recommendations of the Ethical Committee of the University of Essex. Participants were seated in a comfortable armchair in front of a screen during training the BCI and playing a game of "whack a mole". In this game, an image appears on the screen for one second, in which a mole pops its head out of one four holes in the ground (left, right, back, forward). The user is informed of the task before each run. As a result, Conceptual Preparation, and Lemma selection are completed before onset. The user then waits for an auditory cue to covertly speak a command and move a hammer to the correct location to whack the mole. The four chosen covert speech tasks are "BA", "FO", "LE", and "RY", which are Phonetically very dissimilar [196, 214, 217] and can be used as intuitive shortened directional commands. For example, the hammer can hit the left hole by using the covert speech task "LE". For all classes, an identical "Beep" sound was used as the auditory cue. After task presentation, a random rest period between 1 and 3 seconds was placed before the auditory cue to prevent the user from anticipating onset time based on rhythm. This ensures the following linguistic activities begin exactly after auditory cue recognition (trigger driven executive control), and the

system is perfectly synchronized. By using 312ms epochs, the only analyzed Linguistic functions are Phonological Code Retrieval and Syllabification, which are performed automatically by the brain and have an unchangeable duration unique to each individual participant. The experiment has two stages: training and testing. During the training stage, 15 trials of each of the four classes (60 trials in total) are presented to the user in a mixed randomised order. For the testing stage, 5 trials of each of the four classes (20 trials in total) are presented to the user in a mixed randomised order. Figure 5.1 illustrates the experimental protocol.

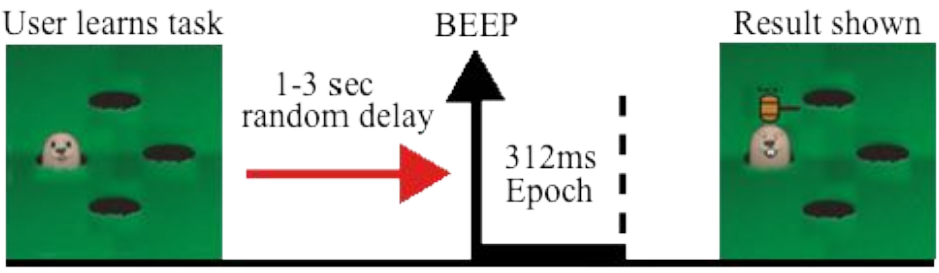



***Figure 5.1:*** *Imagination protocol. An image presented for 1 second shows the mole in one of the holes (left in this case). The user chooses the correct direction for landing the hammer and the correct covert speech task to perform (in this case "LE"). After the task image disappears the user waits to hear the beep sound to begin covertly speaking the correct word. In the testing stage the result is also shown in an image, completing the cycle.*

## *5.2. Data Acquisition and analysis pipeline*

The EEG recording system, data acquisition, preprocessing, feature generation and subset selection, training and testing are identical in this experiment to those for the 312ms trials in chapter four. The only differences are that trials are recorded in randomized order in a single recording run, and that there

are 15 trials per class recorded for training. No cross validation is implemented, and all 15 trials are used at the same time for training. Like chapter four, 4k most valuable Gabor features are identified in the training stage, and vectors of these Gabor features in the 5 online test trials are used for classification.

## *5.3. Classification Accuracy*

The performance of the six participants are presented in Table 5.1. In average, all users got more than 16 out of 20 correct predictions (average performance 82.5%). PLDA was used for classification.

| | Classification Accuracy |
|---|---|
| User 1 | 85% |
| User 2 | 80% |
| User 3 | 75% |
| User 4 | 90% |
| User 5 | 85% |
| User 6 | 80% |
| Mean | 82.5 ± 4.1% |

***Table 5.1:*** *Classification Accuracy for all 6 users.*

## *5.4. Important Times, Frequencies, and Locations*

The outstanding times, frequencies, and brain locations based on the number of instances they have been identified in the most valuable 4K Gabor features, for all 6 users (24K in total) have been presented in figure 5.2. The most important activity begins ~100ms post-onset corresponding to Auditory cue recognition and is concentrated in the 70-128Hz frequency range corresponding to Linguistic functions. Most important regions: Prefrontal Cortex, left STG (Wernicke's area), right, and left IFG (Broca's area). These correspond with Linguistic Phonetic activity prior to articulation.

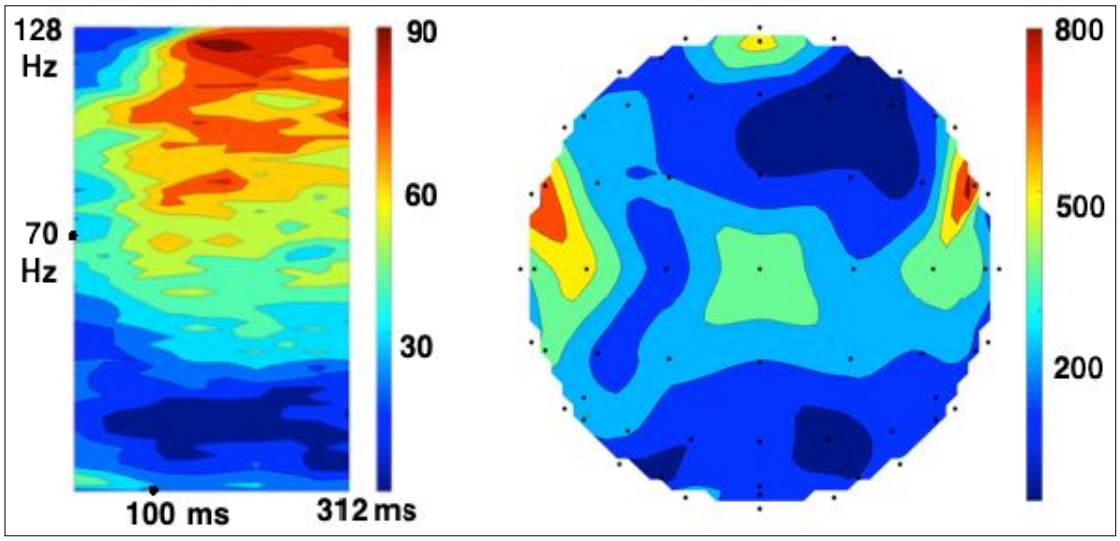



***Figure 5.2: (Left)*** *The cumulative colour-coded joint time-frequency representation of 24K best Gabor features from all users. The most important activity begins ~100ms post-onset corresponding to Auditory cue recognition and is concentrated in the 70-128Hz frequency range corresponding to Linguistic functions.* ***(Right)*** *The topographical plot of the brain using same Gabor features. The top of the plot is the front of the head. Most important regions: Prefrontal Cortex, left STG (Wernicke's area), right, and left IFG (Broca's area). These correspond with Linguistic Phonetic activity prior to articulation.*

## *5.5. Discussion*

As seen in figure 5.2, the key activity begins ~100ms post-onset. This corresponds with the time of Auditory cue recognition. The highest concentration is in the 70-128Hz range, which corresponds with Phonetic Linguistic activities. The most important brain regions are the Prefrontal Cortex (related to trigger-driven executive control, waiting for time cue to begin task), Wernicke's area (related to Phonological code retrieval), and right IFG and Broca's area (related to Syllabification). The spatial, temporal, and spectral properties of the 24K most valuable feature vectors identified from 6 participants, correspond to the automatic linguistic processing stages of word production prior to articulation, and are supported by a substantial body of evidence [23-25, 27, 39-43, 46, 47, 187, 212]. The only conscious effort required from the user to operate this BCI is paying attention and waiting for the Auditory cue to covertly speak the desired command. All brain activities after auditory cue recognition (Phonetic Linguistic functions) are performed automatically by the brain and cannot be modified by the user. In fact, the 312ms trials end before conscious activity of any sort can begin in the brain.

The Linguistic BCI presented in this chapter is just as fast as a traditional P300 system. However, unlike the P300, this novel Linguistic BCI does not require constant gaze and attention to operate correctly. It is impossible to use MI for such a short trial duration.

For operating this novel BCI we used only those Linguistic functions that are completely automatic (Linguistic Phonetic functions) and temporally consistent. It is difficult to maintain perfect consistency for MI tasks, even during the same experiment. Considering these factors, Linguistic BCIs can potentially render MI systems obsolete.

In the present chapter, we only use a fraction of the time-frequency window containing Linguistic activity (up to 170Hz lasting over 2000ms) to successfully classify 4 covert speech tasks. By expanding the time and frequency of analysis, the number of word classes may be increased in future studies. Finally, the performance of the Linguistic BCI may improve by increasing the number of trials used for training [9] with the downside of increasing training duration and greater user fatigue.

To conclude this chapter, plots of the grand average spectrogram for each class from the central channel Cz are presented for all sex users in figures 5.3 to 5.8. Although feature power is not what we use for classification, these plots clearly show differences in each class for all users. Only user 4 perferms over 90 percent (figure 5.6) and the difference between classes in clear in this figure.

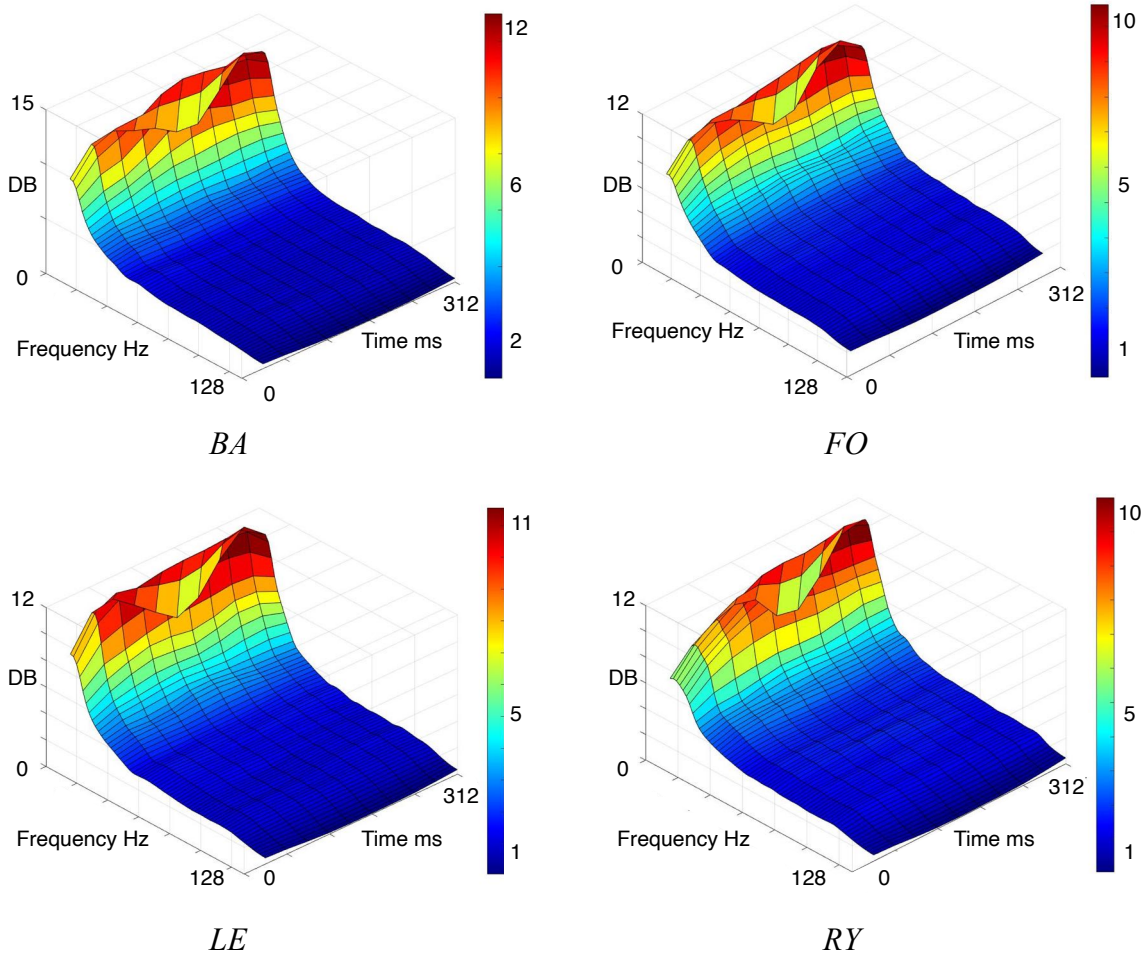



***Figure 5.3:*** *User 1: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

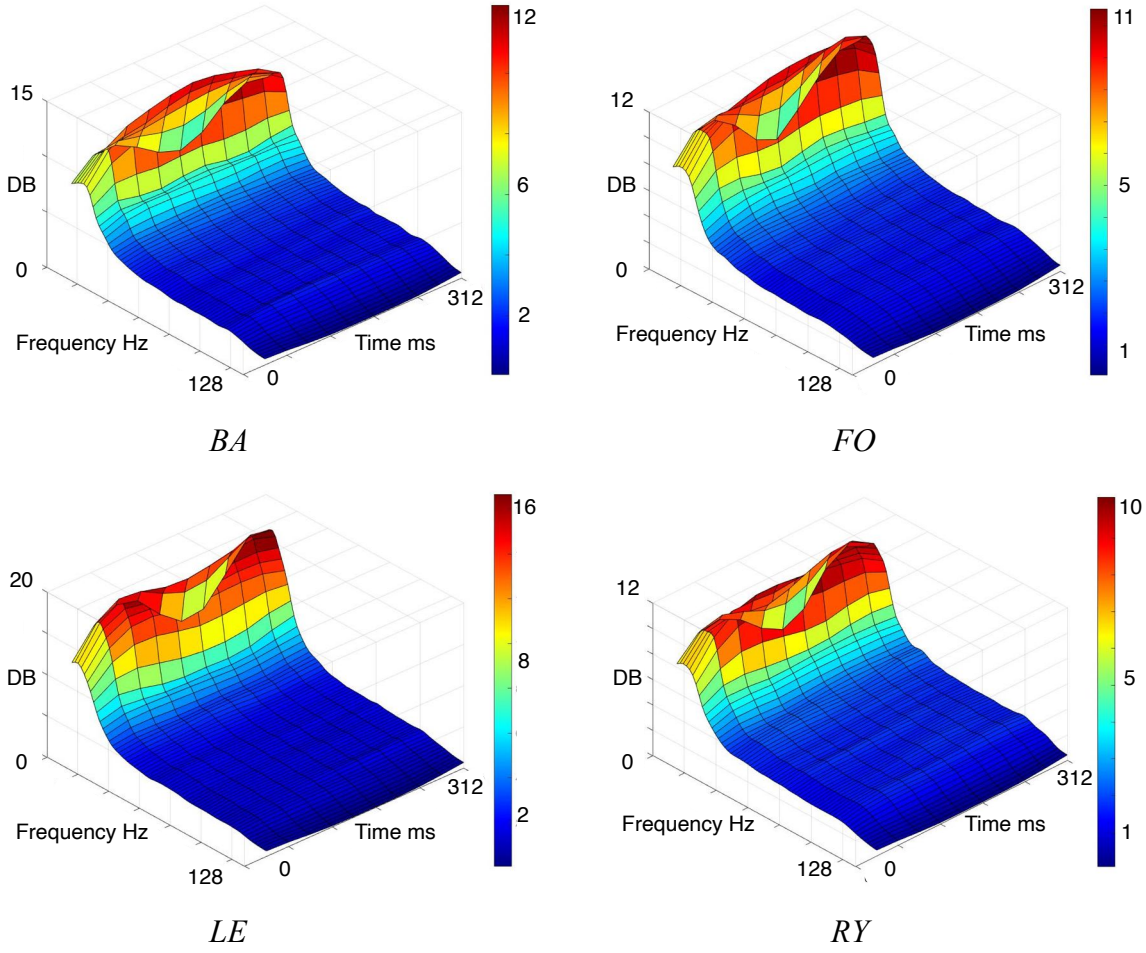



***Figure 5.4:*** *User 2: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

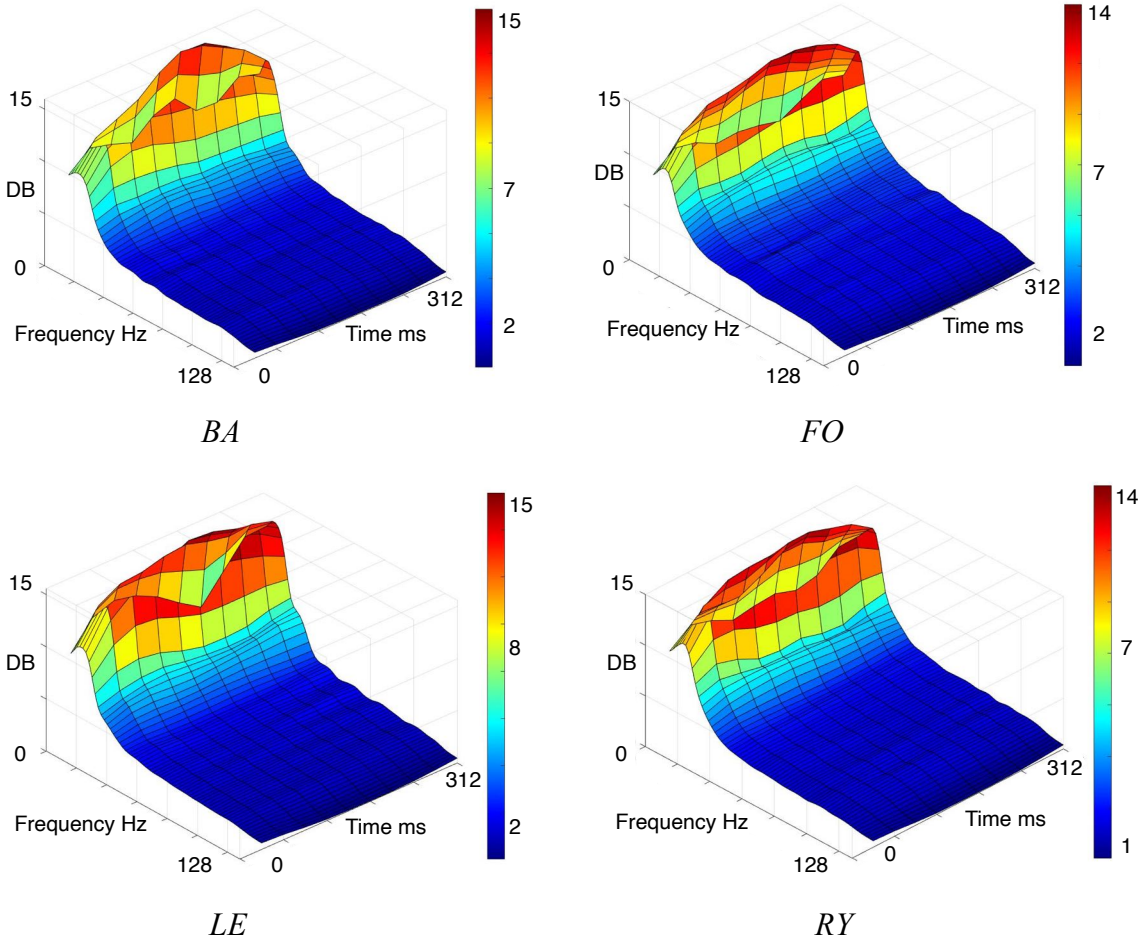



***Figure 5.5:*** *User 3: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

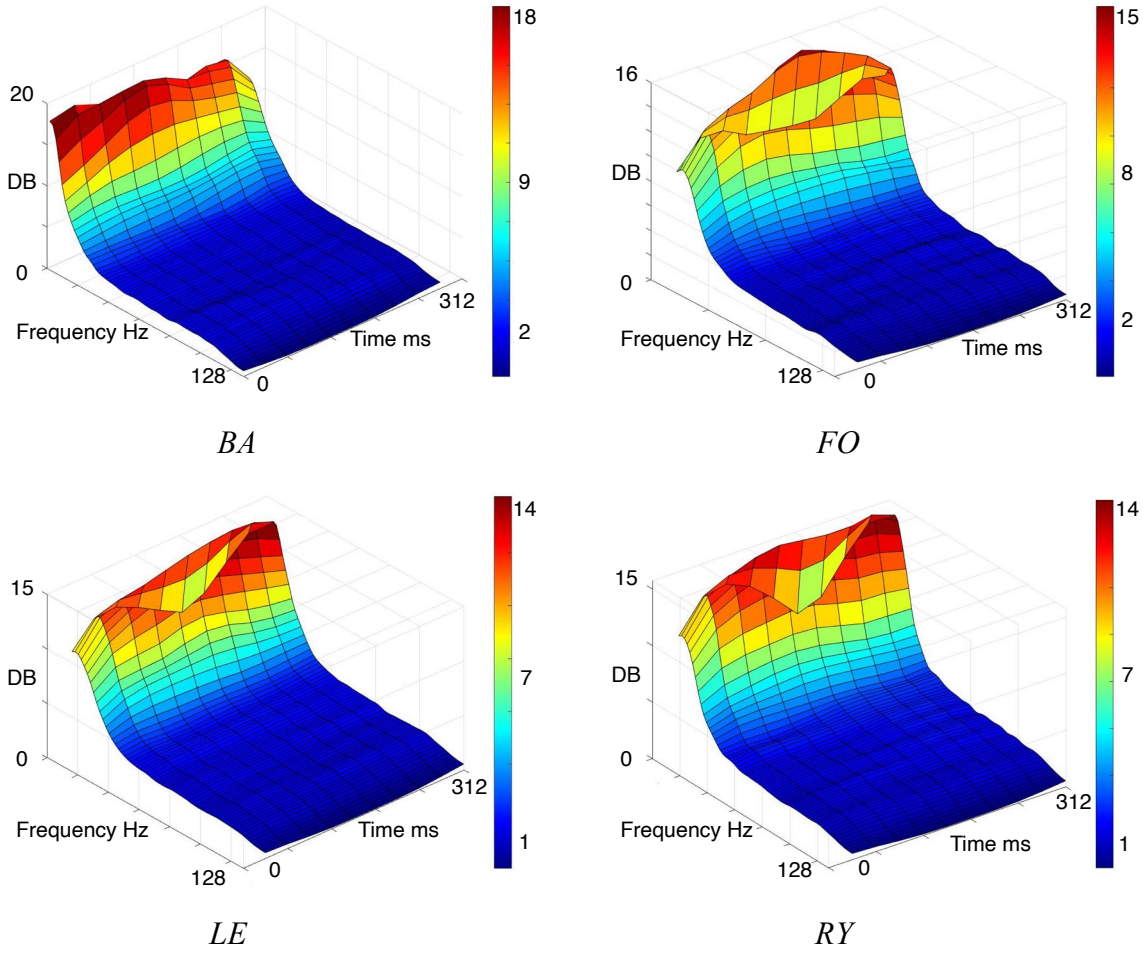



***Figure 5.6:*** *User 4: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

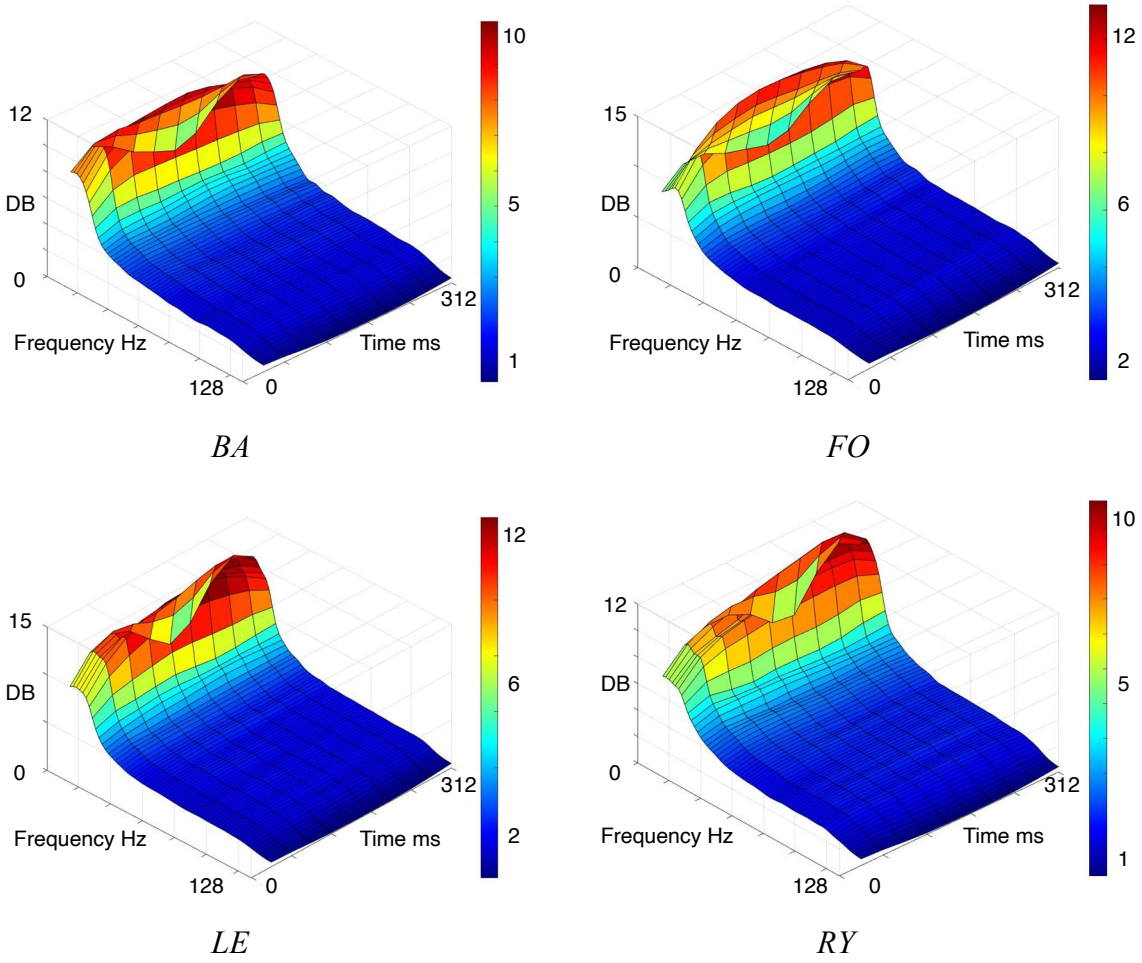



***Figure 5.7:*** *User 5: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

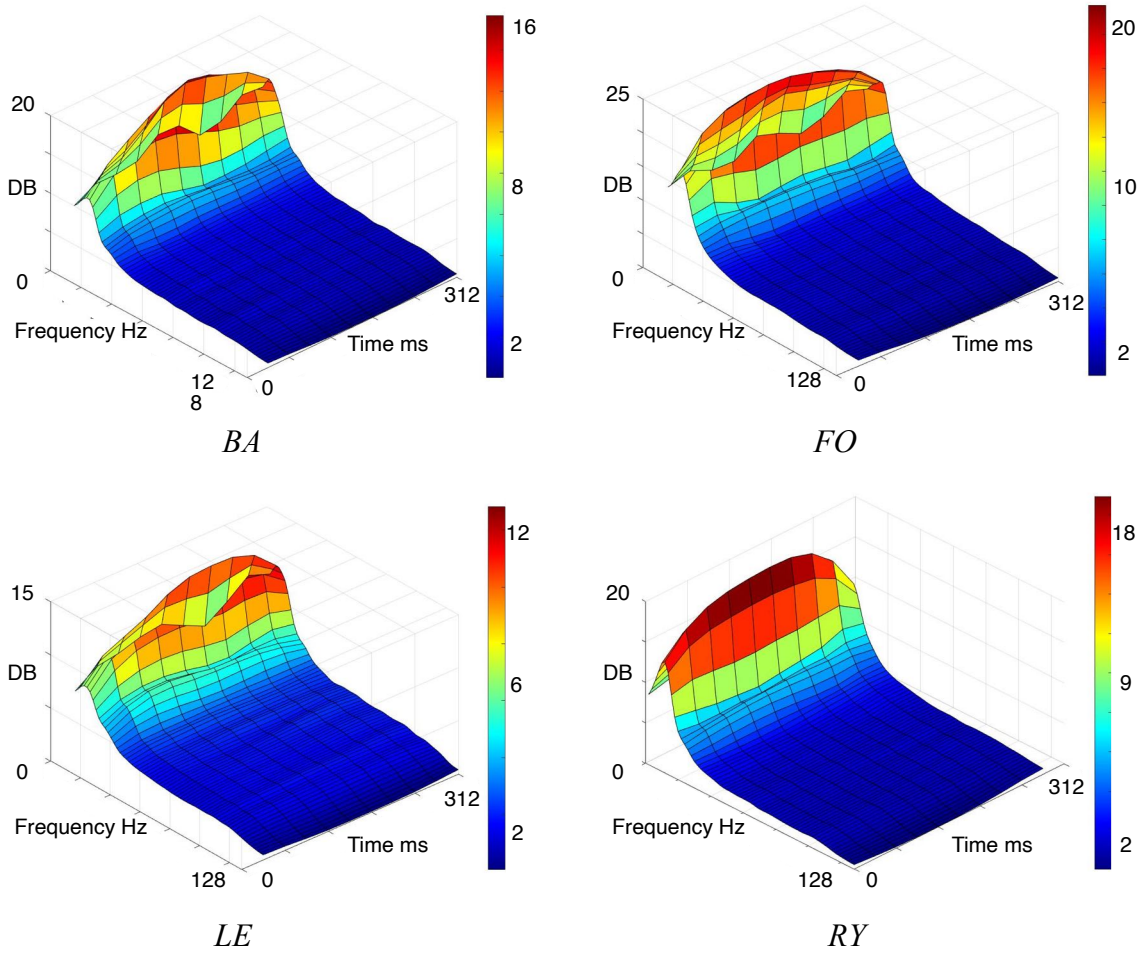
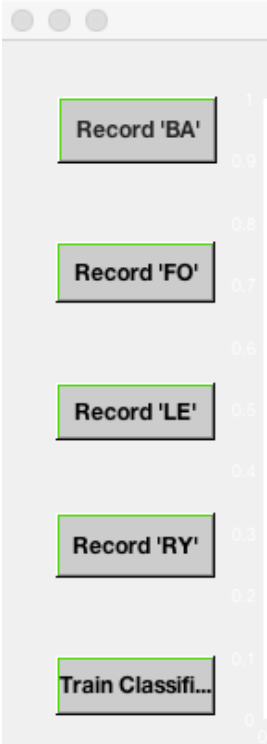



***Figure 5.8:*** *User 6: Grand average spectrogram of each class from all 20 trials (15 training, 5 online testing) and all channels.*

# *Appendix A: Matlab Code*

## *MATLAB ™code for training the BCI*




```
function varargout = Training(varargin)
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
'gui_Singleton',  gui_Singleton, ...
'gui_OpeningFcn', @Training_OpeningFcn, ...
'gui_OutputFcn',  @Training_OutputFcn, ...
'gui_LayoutFcn',  [] , ...
'gui_Callback',   []);
if nargin && ischar(varargin{ 1})
gui_State.gui_Callback = str2func(varargin{1});
end
if nargout
[varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
gui_mainfcn(gui_State, varargin{:});
```

```
end
end
function Training_OpeningFcn(hObject, eventdata, handles, varargin)
handles.output = hObject;
guidata(hObject, handles);
end
function varargout = Training_OutputFcn(hObject, eventdata, handles)
varargout{1} = handles.output;
end
% --- Executes on button press in BA.
function BA_Callback(hObject, eventdata, handles)
t = timer('TimerFcn','ff=0;');
beep on
start(t);
paraport=digitalio('parallel','LPT1');
addline(paraport,0:7,'out');
for s=1:20
putvalue(paraport.line(1),0)
pause(ceil(5*rand (1,1)))
beep
putvalue(paraport.line(1),1)
pause(3)
putvalue(paraport.line(1),0)
end
delete (t);
end
% --- Executes on button press in FO.
function FO_Callback(hObject, eventdata, handles)
t = timer('TimerFcn','ff=0;');
beep on
start(t);
paraport=digitalio('parallel','LPT1');
addline(paraport,0:7,'out');
for s=1:20
putvalue(paraport.line(1),0)
pause(ceil(5*rand (1,1)))
beep
putvalue(paraport.line(1),1)
pause(3)
putvalue(paraport.line(1),0)
end
delete (t);
end
```

```
% --- Executes on button press in LE.
function LE_Callback(hObject, eventdata, handles)
t = timer('TimerFcn','ff=0;');
beep on
start(t);
paraport=digitalio('parallel','LPT1');
addline(paraport,0:7,'out');
for s=1:20
putvalue(paraport.line(1),0)
pause(ceil(5*rand (1,1)))
beep
putvalue(paraport.line(1),1)
pause(3)
putvalue(paraport.line(1),0)
end
delete (t);
end
% --- Executes on button press in RY.
function RY_Callback(hObject, eventdata, handles)
t = timer('TimerFcn','ff=0;');
beep on
start(t);
paraport=digitalio('parallel','LPT1');
addline(paraport,0:7,'out');
for s=1:20
putvalue(paraport.line(1),0)
pause(ceil(5*rand (1,1)))
beep
putvalue(paraport.line(1),1)
pause(3)
putvalue(paraport.line(1),0)
end
delete (t);
end
% --- Executes on button press in Train.
function Train_Callback(hObject, eventdata, handles)
eeglab
ltfatstart
g=gabwin('gauss',8,128,768);
%% Preprocess BA
EEG = pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/BA.bdf',
'channels',[1:64] );
EEG = pop_reref( EEG, []);
```

```
EEG = pop_resample( EEG, 256);
EEG = pop_eegfiltnew(EEG, 49,51,5000,1,[],0);
EEG = pop_epoch( EEG, {  '32767'  }, [0  3], 'newname', 'BDF file resampled
epochs', 'epochinfo', 'yes');
EEG = pop_rmbase( EEG, [0      2996.0938]);
EEG = eeg_checkset( EEG );
pop_export(EEG,'/Users/Amir/Documents/MATLAB/eeglab/Online/BA','elec','
off','time','off','precision',4);
c=zeros(4096,1920);
for CHANNEL=1:64
for TRIAL=1:20
f=BA(CHANNEL,((TRIAL-1)*768+1):1:TRIAL*768);
CC=dgtreal(f,g,8,128);
c(((CHANNEL-1)*64+1):CHANNEL*64,((TRIAL-
1)*96+1):TRIAL*96)=CC(1:64,1:96);
end
end
GC_BA=abs(c);
%% Preprocess FO
EEG = pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/FO.bdf',
'channels',[1:64] );
EEG = pop_reref( EEG, []);
EEG = pop_resample( EEG, 256);
EEG = pop_eegfiltnew(EEG, 49,51,5000,1,[],0);
EEG = pop_epoch( EEG, {  '32767'  }, [0  3], 'newname', 'BDF file resampled
epochs', 'epochinfo', 'yes');
EEG = pop_rmbase( EEG, [0      2996.0938]);
EEG = eeg_checkset( EEG );
pop_export(EEG,'/Users/Amir/Documents/MATLAB/eeglab/Online/FO','elec','
off','time','off','precision',4);
c=zeros(4096,1920);
for CHANNEL=1:64
for TRIAL=1:20
f=FO(CHANNEL,((TRIAL-1)*768+1):1:TRIAL*768);
CC=dgtreal(f,g,8,128);
c(((CHANNEL-1)*64+1):CHANNEL*64,((TRIAL-
1)*96+1):TRIAL*96)=CC(1:64,1:96);
end
end
GC_FO=abs(c);
%% Preprocess LE
EEG = pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/LE.bdf',
'channels',[1:64] );
```

```
EEG = pop_reref( EEG, []);
EEG = pop_resample( EEG, 256);
EEG = pop_eegfiltnew(EEG, 49,51,5000,1,[],0);
EEG = pop_epoch( EEG, {  '32767'  }, [0  3], 'newname', 'BDF file resampled
epochs', 'epochinfo', 'yes');
EEG = pop_rmbase( EEG, [0      2996.0938]);
EEG = eeg_checkset( EEG );
pop_export(EEG,'/Users/Amir/Documents/MATLAB/eeglab/Online/LE','elec','
off','time','off','precision',4);
c=zeros(4096,1920);
for CHANNEL=1:64
for TRIAL=1:20
f=LE(CHANNEL,((TRIAL-1)*768+1):1:TRIAL*768);
CC=dgtreal(f,g,8,128);
c(((CHANNEL-1)*64+1):CHANNEL*64,((TRIAL-
1)*96+1):TRIAL*96)=CC(1:64,1:96);
end
end
GC_LE=abs(c);
%% Preprocess RY
EEG = pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/RY.bdf',
'channels',[1:64] );
EEG = pop_reref( EEG, []);
EEG = pop_resample( EEG, 256);
EEG = pop_eegfiltnew(EEG, 49,51,5000,1,[],0);
EEG = pop_epoch( EEG, {  '32767'  }, [0  3], 'newname', 'BDF file resampled
epochs', 'epochinfo', 'yes');
EEG = pop_rmbase( EEG, [0      2996.0938]);
EEG = eeg_checkset( EEG );
pop_export(EEG,'/Users/Amir/Documents/MATLAB/eeglab/Online/RY','elec','
off','time','off','precision',4);
c=zeros(4096,19200);
for CHANNEL=1:64
for TRIAL=1:20
f=RY(CHANNEL,((TRIAL-1)*768+1):1:TRIAL*768);
CC=dgtreal(f,g,8,128);
c(((CHANNEL-1)*64+1):CHANNEL*64,((TRIAL-
1)*96+1):TRIAL*96)=CC(1:64,1:96);
end
end
GC_RY=abs(c);
%% DBI MATRIX
mean_BA=zeros(4096,96);
```

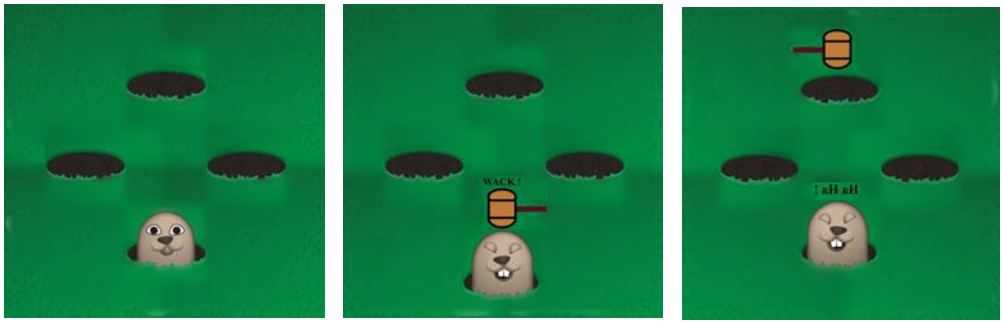
```
std_BA=zeros(4096,96);
mean_FO=zeros(4096,96);
std_FO=zeros(4096,96);
mean_LE=zeros(4096,96);
std_LE=zeros(4096,96);
mean_RY=zeros(4096,96);
std_RY=zeros(4096,96);
for i=1:4096,
for j=1:96,
ba=GC_BA(i,j:96:j+1824);
mean_BA(i,j)=mean(ba);
std_BA(i,j)=std(ba);
fo=GC_FO(i,j:96:j+1824);
mean_FO(i,j)=mean(fo);
std_FO(i,j)=std(fo);
le=GC_LE(i,j:96:j+1824);
mean_LE(i,j)=mean(le);
std_LE(i,j)=std(le);
ry=GC_RY(i,j:96:j+1824);
mean_RY(i,j)=mean(ry);
std_RY(i,j)=std(ry);
end
end
DBI_BA_vs_FO=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_BA_vs_FO(i,j)=(std_BA(i,j)+std_FO(i,j))/(abs(mean_BA(i,j)-
mean_FO(i,j)));
end
end
DBI_BA_vs_LE=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_BA_vs_LE(i,j)=(std_BA(i,j)+std_LE(i,j))/(abs(mean_BA(i,j)-
mean_LE(i,j)));
end
end
DBI_BA_vs_RY=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_BA_vs_RY(i,j)=(std_BA(i,j)+std_RY(i,j))/(abs(mean_BA(i,j)-
mean_RY(i,j)));
end
```

```
end
DBI_LE_vs_FO=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_LE_vs_FO(i,j)=(std_LE(i,j)+std_FO(i,j))/(abs(mean_LE(i,j)-
mean_FO(i,j)));
end
end
DBI_LE_vs_RY=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_LE_vs_RY(i,j)=(std_LE(i,j)+std_RY(i,j))/(abs(mean_LE(i,j)-
mean_RY(i,j)));
end
end
DBI_RY_vs_FO=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_RY_vs_FO(i,j)=(std_RY(i,j)+std_FO(i,j))/(abs(mean_RY(i,j)-
mean_FO(i,j)));
end
end
DBI_ALL=zeros(4096,96);
for i=1:4096,
for j=1:96,
DBI_ALL(i,j)=0.25*(max([DBI_BA_vs_FO(i,j) DBI_BA_vs_LE(i,j)
DBI_BA_vs_RY(i,j)])+max([DBI_BA_vs_FO(i,j) DBI_LE_vs_FO(i,j)
DBI_RY_vs_FO(i,j)])+max([DBI_BA_vs_LE(i,j) DBI_LE_vs_FO(i,j)
DBI_LE_vs_RY(i,j)])+max([DBI_BA_vs_RY(i,j) DBI_RY_vs_FO(i,j)
DBI_LE_vs_RY(i,j)]));
end
end
%% INDEXES OF BEST FEATURES
b=sort(DBI_ALL(:,3),'ascend');
ref=b(4001,1);
sum=0;
indexes=zeros(4000,3);
for i=1:4096,
for j=1:96,
if DBI_ALL(i,j)<ref,
sum=sum+1;
indexes(sum,:)=[i j DBI_ALL(i,j)];
end
```

```
end
end
indexes=sortrows(indexes,3);
indexes=indexes(:,1:2);
fname=['indexes'];
save(fname,'indexes');
%% TRAINING THE CLASSIFICATION OBJECT
num_of_resols=4000;
Training_set=zeros(80,num_of_resols);
y={1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;
2;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;3;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4;4}
;
for n=1:num_of_resols,
i=indexes(n,1);
j=indexes(n,2);
for k=1:20,
p=(k-1)*96;
Training_set(k,n)=GC_BA(i,j+p);
Training_set(k+20,n)=GC_FO(i,j+p);
Training_set(k+40,n)=GC_LE(i,j+p);
Training_set(k+60,n)=GC_RY(i,j+p);
end
end
obj=ClassificationDiscriminant.fit(Training_set,y);
fname=['Classification_Object'];
save(fname,'obj');
%% FILES AND VARIABLES NEEDED FOR ONLINE CODE
percent=0;
I=1;
score=0;
BA=imread ('BA.jpg');
BA_BA=imread ('BA_BA.jpg');
BA_FO=imread ('BA_FO.jpg');
BA_LE=imread ('BA_LE.jpg');
BA_RY=imread ('BA_RY.jpg');
FO=imread ('FO.jpg');
FO_BA=imread ('FO_BA.jpg');
FO_FO=imread ('FO_FO.jpg');
FO_LE=imread ('FO_LE.jpg');
FO_RY=imread ('FO_RY.jpg');
LE=imread ('LE.jpg');
LE_BA=imread ('LE_BA.jpg');
LE_FO=imread ('LE_FO.jpg');
```

```
LE_LE=imread ('LE_LE.jpg');
LE_RY=imread ('LE_RY.jpg');
RY=imread ('RY.jpg');
RY_BA=imread ('RY_BA.jpg');
RY_FO=imread ('RY_FO.jpg');
RY_LE=imread ('RY_LE.jpg');
RY_RY=imread ('RY_RY.jpg');
Blank=imread('Blank.jpg');
[Y,Fs] = audioread('Beep.wav');
[K,FS] = audioread('Cheering.wav');
[J,fS] = audioread('Laugh.wav');
paraport=digitalio('parallel','LPT1');
lines=addline(paraport,0:7,'out');
putvalue(paraport.line(1),0);
putvalue(paraport.line(2),0);
putvalue(paraport.line(3),0);
%%
end
```

## *MATLAB ™ code for testing*



```
function varargout = online(varargin)
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
'gui_Singleton',  gui_Singleton, ...
'gui_OpeningFcn', @online_OpeningFcn, ...
'gui_OutputFcn',  @online_OutputFcn, ...
'gui_LayoutFcn',  [] , ...
'gui_Callback',   []);
if nargin && ischar(varargin{1})
gui_State.gui_Callback = str2func(varargin{1});
end
if nargout
[varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
gui_mainfcn(gui_State, varargin{:});
end
function online_OpeningFcn(hObject, eventdata, handles, varargin)
handles.output = hObject;
guidata(hObject, handles);
function varargout = online_OutputFcn(hObject, eventdata, handles)
varargout{1} = handles.output;
end
% --- Executes on button press in pushbutton1.
function pushbutton1_Callback(hObject, eventdata, handles)
function start_message(~,~)
axis([-10 10 -10 10]);
t=text(-7,0, 'your score is
',num2str(percent),'%','FontSize',50,'Color','red','Visible','on');
end
function next_message(~,~)
t.Visible='off';
u=text(-7,0,'Prepare for next task','FontSize',50,'Color','red','Visible','on');
delete (T1);
```

```
end
function first_image(~,~)
u.Visible='off';
if target==1
imshow(BA);
end
if target==2
imshow(FO);
end
if target==3
imshow(LE);
end
if target==4
imshow(RY);
end
delete (T2);
end
function blank_image(~,~)
imshow(Blank);
delete (T3);
end
function start_recording(~,~)
putvalue(paraport.line(2),1);
delete (T4);
end
function send_trigger(~,~)
putvalue(paraport.line(1),1);
beep
delete (T5);
end
function stop_recording_classify(~,~)
putvalue(paraport.line(3),1);
% preprocess data
EEG =
pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/TEST_IN.bdf',
'channels',[1:64] );
EEG = pop_reref( EEG, []);
EEG = pop_resample( EEG, 256);
EEG = pop_eegfiltnew(EEG, 49,51,5000,1,[],0);
EEG = pop_epoch( EEG, {  '32767'  }, [0  3], 'newname', 'BDF file resampled
epochs', 'epochinfo', 'yes');
EEG = pop_rmbase( EEG, [0      2996.0938]);
EEG = eeg_checkset( EEG );
```

```
pop_export(EEG,'/Users/Amir/Documents/MATLAB/eeglab/Online/TEST','elec','off','time','off','precision',4);
% calculate GC_IN matrix
c=zeros(4096,96);
for CHANNEL=1:64
f=TEST(CHANNEL,:);
CC=dgtreal(f,g,8,128);
c(((CHANNEL-1)*64+1):CHANNEL*64,1:96)=CC(1:64,1:96);
end
GC_IN=abs(c);
% classify
test_data=zeros(1,4000);
for a=1:4000,
i=indexes(a,1);
j=indexes(a,2);
test_data(1,a)=GC_IN(i,j);
end
label=predict(obj,test_data);
delete (T6);
end
%
function give_feedback(~,~)
if target==1
if label==1
score=score+1;
sound(K,FS);
imshow(BA_BA)
end
if label==2
imshow(BA_FO)
end
if label==3
imshow(BA_LE)
end
if label==4
imshow(BA_RY)
end
end
if target==2
if label==1
imshow(FO_BA)
end
if label==2
```

```
score=score+1;
sound(K,FS);
imshow(FO_FO)
end
if label==3
imshow(FO_LE)
end
if label==4
imshow(FO_RY)
end
end
if target==3
if label==1
imshow(LE_BA)
end
if label==2
imshow(LE_FO)
end
if label==3
imshow(LE_LE)
sound(K,FS);
score=score+1;
end
if label==4
imshow(LE_RY)
end
end
if target==4
if label==1
imshow(RY_BA)
end
if label==2
imshow(RY_FO)
end
if label==3
imshow(RY_LE)
end
if label==4
imshow(RY_RY)
sound(K,FS);
score=score+1;
end
end
```

```
%calculate percent
percent=score/I*100;
delete (T7);
end
function save_reset(~,~)
imshow(Blank);
%reset lines
putvalue(paraport.line(1),0);
putvalue(paraport.line(2),0);
putvalue(paraport.line(3),0);
%save BDF file for next trial
EEG =
pop_biosig('/Users/Amir/Documents/MATLAB/eeglab/Online/SHORT_FILE.b
df');
EEG = eeg_checkset( EEG );
pop_writeeeg(EEG,
'/Users/Amir/Documents/MATLAB/eeglab/Online/TEST_IN.bdf',
'TYPE','BDF');
%save GC_IN
fname=['GC_IN_',num2str(I)];
save(fname,'GC_IN','target','label');
delete (T8);
delete (T9);
end
% The sequence of task which are presented to the user
task=[1 2 3 4 2 3 4 1 3 4 1 2 4 1 2 3 4 3 2 1 1 4 3 2 2 1 4 3 3 2 1 4 2 3 4 1 3 4 1
2];
for I=1:40
target=task(I);
T1 = timer('TimerFcn',@start_message);
T2 = timer('TimerFcn',@next_message,'StartDelay',2);
T3 = timer('TimerFcn',@first_image,'StartDelay',4);
T4 = timer('TimerFcn',@blank_image,'StartDelay',7);
T5 = timer('TimerFcn',@start_recording,'StartDelay',7.9);
T6 = timer('TimerFcn',@send_trigger,'StartDelay',8);
T7 = timer('TimerFcn',@stop_recording_classify,'StartDelay',11.1);
T8 = timer('TimerFcn',@give_feedback,'StartDelay',13);
T9 = timer('TimerFcn',@save_reset,'StartDelay',15);
start([T1 T2 T3 T4 T5 T6 T7 T8 T9]);
end
end
end
end
```

# Appendix B: Hardware setup, Data acquisition, Pre-processing

## *B.1. Hardware and software setup for the BioSemi system*

The EEG signals were recorded using a 64 channel "Biosemi Active-Two" system [76]. The user is seated comfortably in front of a large screen approximately two meters away. A cap is fitted onto the user's head, and 64 electrodes are connected to the cap according to the 10-10 system. The sampling frequency is initially set to 2048 Hz. Figure B.1 shows the setup of the laboratory.

***Figure B.1:.*** *Brain-Computer-Interface laboratory, School of Computer Science and Electronic Engineering, University of Essex.*

The Active-Two system requires trigger signals via a parallel port. These triggers are visible in the recorded EEG data and can be used to identify the exact

moment the user is presented with a queue. To avoid problems with synchronization, the triggers are generated by the same computer (clock) that runs the graphical user interface. The GUI was created with MATLAB™ and the same section of the code initiates the trigger signal for the ActiveTwo, and the cue for the user. In the case of visual cues, the desired word will appear on the screen in front of the user. The word will disappear after one second. The user is asked to perform the covert speech task as soon as they see the word on the screen. Figure B.2 illustrates the GUI and what the user sees. The command appears for one second and disappears before onset.


***Figure B.2:*** *The MATLAB™ generated graphical user interface showing the next task before screen goes blank and audio cue is generated.*

A second computer records the EEG signals. This computer is connected to the Active-Two system via USB port. The BioSemi acquisition software (ActiView) is also installed on this computer. The ActiveTwo has a pre-amplifier stage on the electrode and can correct for high impedances. However, the offset

voltage between the A/D box and the body was kept between 25mV and 50mV as recommended by the manufacturer.

## *B.2 Data acquisition and pre-processing*

The pre-processing was done with the use of the EEGLAB package [122], a MATLAB™ toolbox created by BioSemi. The recorded EEG data is saved by the BioSemi software in a format called "BDF". These files contain the raw data for at least 64 channels, at a sampling rate of 2048Hz. So, the first step is to select only the 64 channels, which are connected to the user's scalp. Figure B.3 shows the dialogue box generated by EEGLAB.

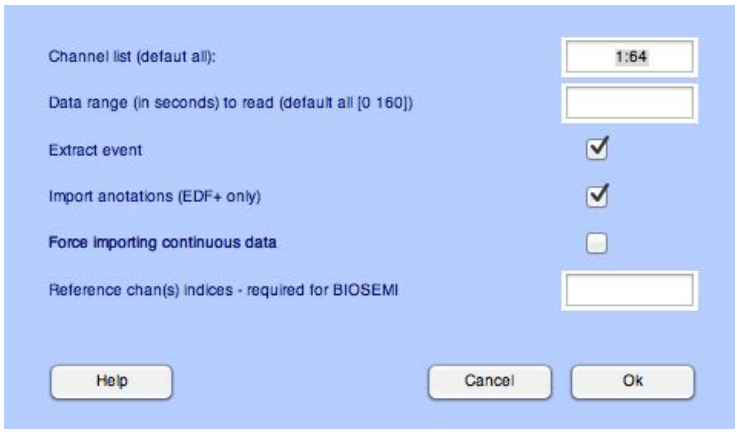



***Figure B.3:*** *selecting only the first 64 channels from all possible 256.*

After the correct channels have been selected, the first thing to do is to reduce the sampling rate. A good sampling frequency would be 256 Hz, providing

a frequency content of 1-128 Hz for the data. Figure B.4 shows the dialogue box generated by EEGLAB to adjust sampling rate.

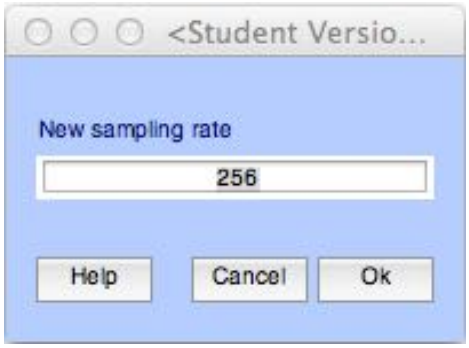



***Figure B.4:*** *adjusting the sampling rate from 2048 to 256Hz.*

EEG data is in the form of sampled voltage from the probe electrodes. To analyse these values, the voltages need a reference point. In the measurement and recording stage two extra electrodes provide the ground. However, for the signal processing stage, an average reference is selected. Figure B.5 illustrates the dialogue box generated by EEGLAB for computing average references.

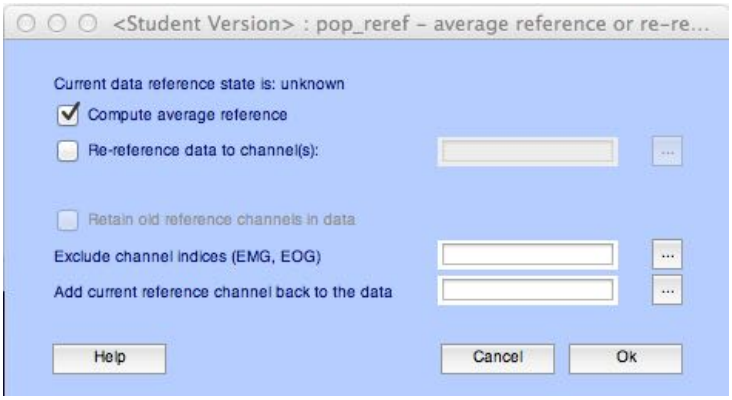



***Figure B.5:*** *The options for re-referencing the data in EEGLAB*

The next step is to filter the data. The frequency range I selected to study is 5-100Hz. Although, there are suggestions that there may be useful information in EEG data in higher frequencies. If the results do not show promise, the frequency range may be increased beyond 100 Hz. Figure B.6 illustrated the use of a finite impulse response filter of order 100 to select the (5-100) Hz band.

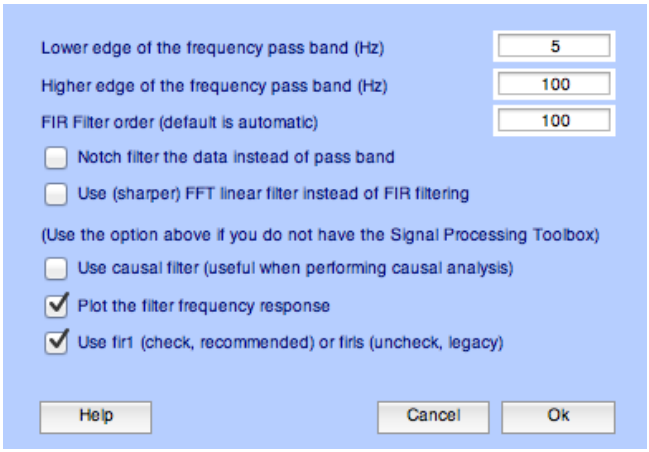



***Figure B.6:*** *FIR order 100 filter to select the 5-100 Hz range.*

As seen in figure B.7, the filter amplitude response has a sharp cut-off with no ripples in the selected data range and this is the reason for selecting this filter.

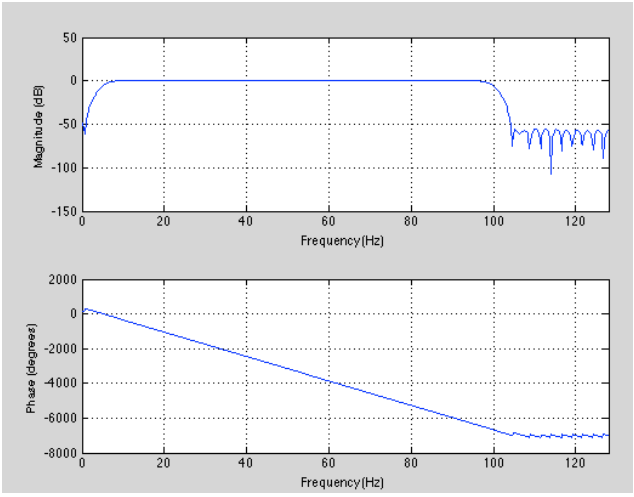



***Figure B.7:*** *The magnitude and phase response of the FIR filter.*

The electricity cables in the lab generate 50 Hz noise (in the UK). This noise has greater amplitude than the EEG signals and could lead to errors in later computations. The next step is to notch filter the (49-51) Hz band out of the signal with a FIR order 100 filter. As before, the acceptable performance is the reason for selecting this type and order of filter. Figure B.8 illustrates the setup of the dialogue box in EEGLAB.

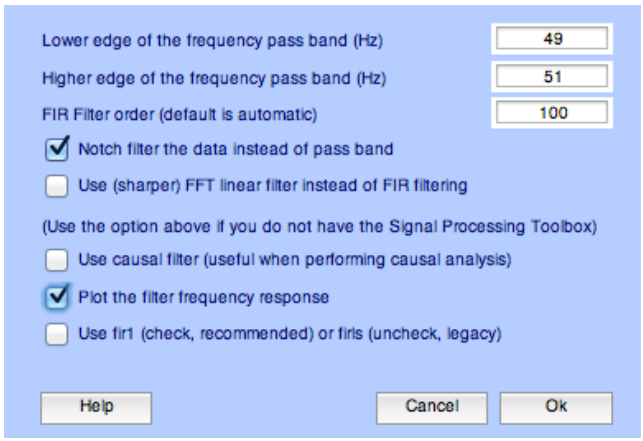



***Figure B.8:*** *Setting a FIR order 100 notch filter for (49-51) Hz.*

Figure B.9 presents the magnitude and phase response of the notch filter.

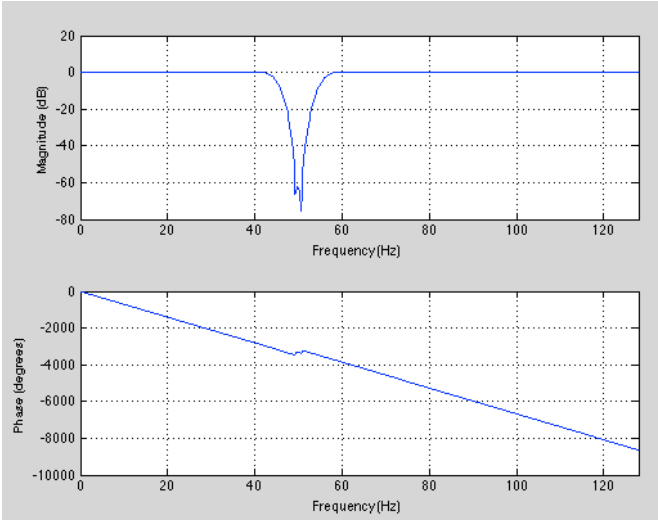



***Figure B.9:*** *Magnitude and phase response of the FIR notch filter.*

EOG and EMG artifacts were removed by the Automatic Artifact Rejection (AAR) utility in EEGLAB [123]. In cases where EOG electrodes are not used, the AAR toolbox uses Blind Source Separation [198] for decomposition of correlative EEG frames with sliding windows. The default BSS algorithm for EOG removal, is the "Second Order Blind Identification" or SOBI [218, 219]. In this case, the EEG can be found by minimizing cross-correlations and maximizing auto-correlations. EOG artifacts are identified by their small Fractal Dimension [198] and removed. The default setup, for removing EOG using the AAR toolbox id presented in figure B.10.

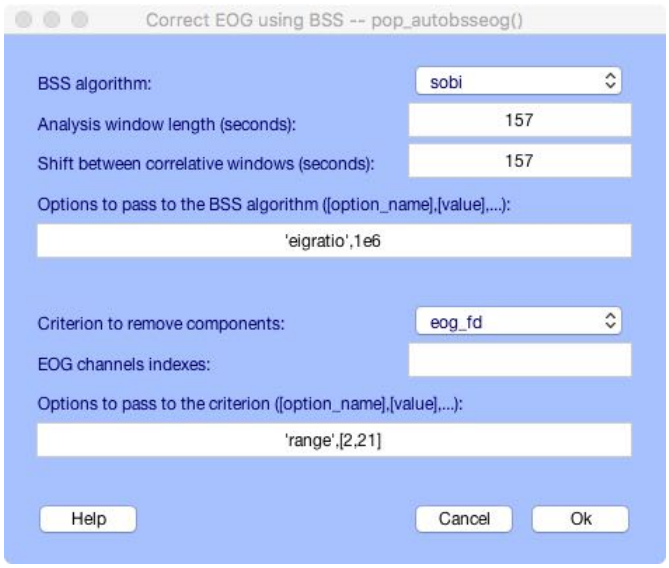



***Figure B.10:*** *EEG decomposition with SOBI [219] and removing EOG artefacts by FD [198].*

EMG artifacts cover a broad frequency range, and like white noise, have lower autocorrelation compared to EEG signals. For elimination of EMG artefacts, the default BSS algorithm is Canonical Correlation Analysis [199] or

CCA. EMG artefacts are identified by their low spectral power density and removed. Figure B.11 Presents the default setup in the AAR toolbox.

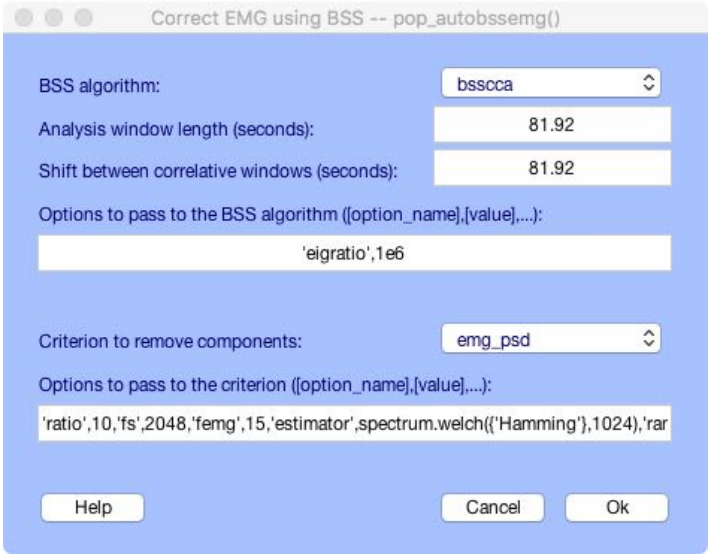



***Figure B.11:*** *EEG decomposition with CCA [199] and removing EMG artefacts PSD [123].*

Now it is time to epoch the data. From each of the 5 recording runs, 12 trials with a 2 second duration are selected for each word. Word-specific time locking events, which are visible as vertical lines over the EEG signals are generated at the exact instant the user is presented with a cue. Figure B.12 illustrates the dialogue box in EEGLAB for selecting epochs.

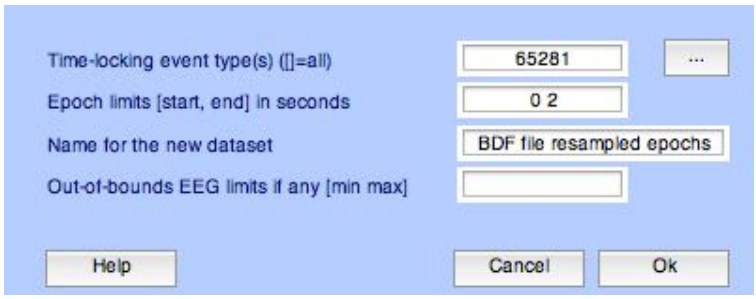



***Figure B.12:*** *Selecting trials of length 2 seconds, using the word-specific time locking event.*

One of the best software platforms for signal processing is MATLAB™. However, EEGLAB does not provide a direct export facility to MATLAB™ to save the data in the form of a 64x15360 matrix. The data is exported to a text file, and then converted into a matrix with the use of MATLAB's data import facility. Figure B.13 demonstrates the setup of the EEGLAB™ export tool. The data is organized correctly and there is no need to add time stamps and channel stamps. The location of each sample contains this information. The number of significant digits is chosen to be 4. Figure B.14 shows MATLAB's data import tool.

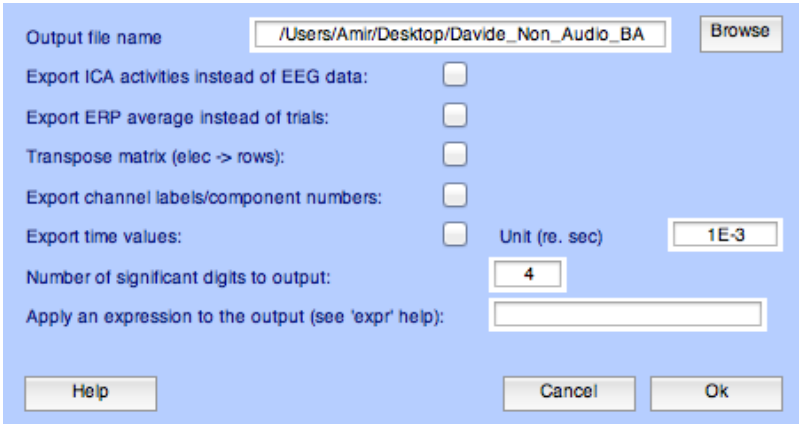



***Figure B.13:*** *Exporting the data to a text file by EEGLAB.*

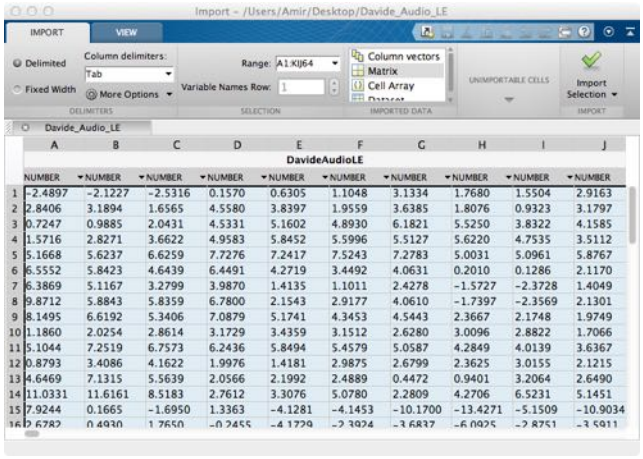



***Figure B.14:*** *Importing data from text file and saving it as a numeric matrix in MATLAB™.*

# *Bibliography*

[1] J. McFadden, "The conscious Electromagnetic information field theory," *Journal of Consciousness Studies,* vol. 9, no. 8, pp. 45-60, 2002.

[2] J. McFadden, "The CEMI field theory, Seven clues to the nature of consciousness," in *The Emerging Physics of Consciousness*, J. A. Tuszynski Ed. Heidelberg: Springer Berlin, 2006, ch. 12, pp. 385-404.

[3] V. Astakhov, "Continuum of consciousness, mind uploading and resurrection of human consciousness," in *conference Toward a Science of Consciousness*, April 8-12 2008.

[4] M. B. Mensky and P. N. Lebedev, "Quantum features of consciousness, computers and brain," 2009.

[5] K. Watkins and T. Paus, "Modulation of Motor Excitability during Speech Perception: The Role of Broca's Area," *Cognitive Neuroscience,* vol. 16, no. 6, pp. 978-987, 2004.

[6] M. Heiser, M. Lacoboni, F. Maeda, J. Markus, and J. C. Mazziotta, "Short communication the essential role of Broca's area in imitation," *European journal of Neuroscience,* vol. 17, pp. 1123-1128, 2003.

[7] M. DZmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward EEG Sensing of Imagined Speech," in *13th Conference, HCI International*, 2009, pp. 40-48, doi: 10.1007/978-3-642-02574-7_5.

[8] K. Brigham and V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *Biometrics: Theory Applications and Systems (BTAS)*, 27-29 Sept. 2010 2010, pp. 1-8, doi: 10.1109/BTAS.2010.5634515.

[9] E. Kaan, "Event related potentials and language processing: a brief overview," *Language and linguistic compass,* vol. 1, no. 6, pp. 571-579, 2007.

[10] L. Y. Ganushchak, I. K. Christoffels, and N. O. Schiller, "The use of electroencephalography in language production research: a review," *Frontiers in psychology,* vol. 2, p. 208, 2011, doi: 10.3389/fpsyg.2011.00208.

[11] C. Cooney, Rafaella Folli, and Damien Coyle, "Mel Frequency Cepstral Coefficients Enhance Imagined Speech Decoding Accuracy from EEG," presented at the *9th Irish Signals and Systems Conference (ISSC*, 2018.

[12] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features.," *Journal of neural engineering,* vol. 15, no. 1, 2018.

[13] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural networks : the official journal of the International Neural Network Society,* vol. 22, no. 9, pp. 1334-9, Nov 2009, doi: 10.1016/j.neunet.2009.05.008.

[14] H. Gurkok, M. Poel, and J. Zwiers, "Classifying motor imagery in presence of speech," in *Neural Networks (IJCNN)*, Barcelona, 2010, pp. 1098-7576, doi: 10.1109/IJCNN.2010.5595733.

[15] S. Iqbal, Y. U. Khan, and O. Farooq, "EEG based classification of imagined vowel sounds," in *Computing for Sustainable Global Development, 2nd International Conference on*, 2015, pp. 1591-1594.

[16] K. Brigham and V. Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication A Preliminary Investigation into Synthetic Telepathy," in *Bioinformatics and Biomedical Engineering (iCBBE)*, Chengdu, 18-20 June 2010 2010, pp. 1-4, doi: 10.1109/ICBBE.2010.5515807.

[17] L. Wang, X. Zhang, and Y. Zhang, "Extending motor imagery by speech imagery for brain-computer interface," in *35th Annual International Conference of the IEEE EMBS*, 2013, pp. 7056-7059.

[18] M. Matsumoto and J. Hori, "Classification of silent speech using adaptive collection," in *Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT)*, 16-19 April 2013 2013, pp. 5-12, doi: 10.1109/CIRAT.2013.6613816.

[19] M. Matsumoto and J. Hori, "Classification of silent speech using support vector machine and relevance vector machine," *Applied Soft Computing,* vol. 20, pp. 95-102, 2014, doi: 10.1016/j.asoc.2013.10.023.

[20] A. Riaz, S. Akhtar, S. Iftikhar, A. A Khan, and A. Salman, "Inter comparison of classification techniques for vowel speech imagery using EEG sensors," in *Systems and Informatics (ICSAI)*, 15-17 Nov. 2014 2014, pp. 712-717, doi: 10.1109/ICSAI.2014.7009378.

[21] K. Taekyung, L. Jeyeon, C. Hoseok, H. L, K. In-Young, and J. D. Pyo, "Meaning based covert speech classification for brain-computer interface based on electroencephalography," in *Neural Engineering (NER)*, 6-8 Nov. 2013 2013, pp. 53-56, doi: 10.1109/NER.2013.6695869.

[22] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis and classification of speech imagery EEG for BCI," *Biomedical Signal Processing and Control,* vol. 8, no. 6, pp. 901-908, 2013, doi: 10.1016/j.bspc.2013.07.011.

[23] J. D. Greenlee *et al.*, "Human Auditory Cortical Activation during Self-Vocalization," *PloS one,* vol. 6, no. 3, pp. 1-15, 2011.

[24] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," *Frontiers in neuroscience,* vol. 5, p. 65, 2011, doi: 10.3389/fnins.2011.00065.

[25] A. Llorens, A. Trebuchon, C. Liegeois-Chauvel, and F. X. Alario, "Intra-cranial recordings of brain activity during language production," *Frontiers in psychology,* vol. 2, p. 375, 2011, doi: 10.3389/fpsyg.2011.00375.

[26] B. N. Pasley *et al.*, "Reconstructing speech from human auditory cortex," *PLoS biology,* Research Support, N.I.H., Extramural vol. 10, no. 1, p. e1001251, Jan 2012, doi: 10.1371/journal.pbio.1001251.

[27] P. Indefrey and W. J. M. Levelt, "The spatial and temporal signatures of word production components," *Cognition,* vol. 92, pp. 101-144, 2004.

[28] M. Steinschneider, K. V. Nourski, H. Kawasaki, H. Oya, J. F. Brugge, and M. A. Howard, 3rd, "Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus," *Cerebral cortex,* vol. 21, no. 10, pp. 2332-47, Oct 2011, doi: 10.1093/cercor/bhr014.

[29] X. Tian and D. Poeppel, "Mental imagery of speech and movement implicates the dynamics of internal forward models," *Frontiers in psychology,* vol. 1, p. 166, 2010, doi: 10.3389/fpsyg.2010.00166.

[30] X. Tian and D. Poeppel, "Mental imagery of speech: linking motor and pereptual systems through internal simulation and estimation," *Frontiers in human neuroscience,* vol. 2, 2012.

[31] A. B. Benevides, T. F. Bastos, and M. Sarcinelli Filho, "proposal of BCI architecture to command a robotic wheelchair," in *IEEE International Symposium on Industrial Electronics*, Gdansk, 2011: IEEE, pp. 2249-54, doi: 10.1109/ISIE.2011.5984511.

[32] A. Toyomura, K. Suzuki, K.Yokosawa, and S. Kuriki, "Prediction of vocalization using time-frequency information of mu rhythm," in *SICE Annual Conference 2010, Proceedings of*, 2010: IEEE, pp. 3407–8.

[33] S. Martin, 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,* vol. 12, 2018.

[34] E. C. Leuthardt *et al.*, "Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task," *Frontiers in human neuroscience,* vol. 6, p. 99, 2012, doi: 10.3389/fnhum.2012.00099.

[35] M. A. Pitts, J. Padwal, D. Fennelly, A. Martínez, and S. A. Hillyard, "Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness," *NeuroImage,* vol. 101, pp. 337-350, 2014.

[36] F. Darvas, R. Scherer, J. G. Ojemann, R..P. Rao, K. J. Miller, and L. B. Sorensen, "High gamma mapping using EEG," *Neuroimage,* vol. 49, no. 1, pp. 930-938, 2010.

[37] S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations," *Front. Hum. Neurosci.,* vol. 7, no. 138, 2013.

[38] J. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Frontiers in Human Neuroscience,* vol. 3, p. 61, 2009.

[39] P. Indefrey, "The spatial and temporal signatures of word production components: a critical update," *Frontiers in psychology,* vol. 2, no. 255, pp. 255-271, 2011.

[40] M. Fukuda, R. Rothermel, C. Juhász, M. Nishida, S. Sood, and E. Asano, "Cortical gamma-oscillations modulated by listening and overt repetition of phonemes," *NeuroImage,* vol. 49, no. 3, pp. 2735–2745, 2010.

[41] V. L.Towle *et al.*, "ECoG gamma activity during a language task: differentiating expressive and receptive speech areas," *Brain : a journal of neurology,* vol. 131, pp. 2013-2027, 2008.

[42] A. Flinker, E. F. Chang, H. E. Kirsch, N. M. Barbaro, N. E. Crone, and R. T. Knight, "Single-trial speech suppression of auditory cortex activity in humans," *The Journal of neuroscience : the official journal of the Society for Neuroscience,* vol. 30, no. 49, pp. 16643-50, Dec 8 2010, doi: 10.1523/JNEUROSCI.1809-10.2010.

[43] O. Creutzfeldt, G. Ojemann, and E. Lettich, "Neuronal activity in the human lateral temporal lobe," *Experimental BrainResearch,* vol. 77, pp. 451-475, 1989.

[44] R.. C. Martin, M. F. Lesch, and M. C. Bartha, "Independence of Input and Output Phonology in Word Processing and Short-Term Memory," *Journal of Memory and Language,* vol. 41, pp. 3-29, 1999.

[45] M. Leinenger, "Phonological coding during reading," *Psychol Bull,* vol. 140, no. 6, pp. 1534-55, Nov 2014, doi: 10.1037/a0037830.

[46] J. Numminena and G. Curio, "Differential effects of overt, covert and replayed speech on vowel- evoked responses of the human auditory cortex," *Neuroscience Letters,* vol. 272, pp. 29-32, 1999.

[47] S. Chakrabarti, H. M. Sandberg, J. S. Brumberg, and D. J. Krusienski, "Progress in Speech Decoding from the Electrocorticogram," *Biomed Eng Lett,* vol. 5, pp. 10-21, 2015.

[48] X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw, and G. Schalk, "Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition," *NeuroImage,* vol. 54, pp. 2960-2972, 2011.

[49] B. Graimann, G. Pfurtscheller, and B. Allison, *Brain-computer interfaces, Applying our Minds to Human-Computer Interaction*. Springer, 2010, p. 408.

[50] J. Perkel and F. Guenther, "The sensory motor control of speech production," in *the proceedings of the international symposium on measurement, analysis, and modeling of human functions*, 2001, pp. 359-365.

[51] R. Carter, *The human brain, an illustrated guide to its structure, function, and disorders*. DK, 2009.

[52] J. N. Mak and J. R. Wolpaw, "Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects," *IEEE Rev Biomed Eng.,* vol. 2, pp. 187-199, 2009.

[53] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain computer interfaces for communication and control," *Clinical Neurophysiology,* vol. 113, pp. 767–791, 2002.

[54] M. van Gerven *et al.*, "The brain-computer interface cycle," *Journal of neural engineering,* Research Support, Non-U.S. Gov't
Review vol. 6, no. 4, p. 041001, Aug 2009, doi: 10.1088/1741-2560/6/4/041001.

[55] R. E. Isaacs, D. J. Weber, and A. B. Schwartz, "Work Toward Real-Time Control of a Cortical Neural Prosthesis," *IEEE Transactions on Rehabilitation Engineering,* vol. 8, no. 2, pp. 196 - 198, 2002.

[56] E. C. Leuthardt *et al.*, "Using the electrocorticographic speech network to control a brain–computer interface in humans," *Neural Eng,* vol. 8, no. 3, p. 11, 2011.

[57] B. Litt, "Implantable devices for epilepsy: a clinical perspective," in *Engineering in Medicine and Biology*, 2002: IEEE, pp. 2035-36.

[58] S. Senay, L. F. Chaparro, M. Sun, and R. J. Sclabassi, "Asynchronous Sigma Delta Modulator based subdural neural implants for epilepsy patients," in *IEEE 35th Annual Northeast Bioengineering Conference*, 2009: IEEE, pp. 1-2, doi: 10.1109/NEBC.2009.4967640.

[59] J. A. Wilson, E. A. Felton, P. C. Garell, G. Schalk, and J. C. Williams, "ECoG factors underlying multimodal control of a brain-computer interface," *IEEE Trans Neural Syst Rehabil Eng,* vol. 14, no. 2, pp. 246-250, 2006.

[60] Xi. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," *J. Neural Eng.,* vol. 8, 2011.

[61] S. Ikeda, et al., "Neural decoding of single vowels during covert articulation using electrocorticography," *Frontiers in Human Neuroscience,* vol. 8, 2014.

[62] S. Martin, et al, "Decoding spectrotemporal features of overt and covert speech from the human cortex," *Frontiers in neuroengineering,* vol. 7, 2014.

[63] S. Martin, et al., "Word pair classification during imagined speech using direct brain recordings," *Scientific reports,* vol. 6, 2016.

[64] K. Ibayashi, Naoto Kunii, Takeshi Matsuo, Yohei Ishishita, Seijiro Shimada, Kensuke Kawai, and Nobuhito Saito, "Decoding Speech With Integrated Hybrid Signals Recorded From the Human Ventral Motor Cortex," *Frontiers in neuroscience,* vol. 12, 2018.

[65] Q. Rabbani, Griffin Milsap, and Nathan E. Crone, "The Potential for a Speech Brain–Computer Interface Using Chronic Electrocorticography," *Neurotherapeutics,* vol. 16, no. 1, pp. 144-165, 2019.

[66] J. Kohlmorgen and B. Blankertz, "Bayesian classification of single-trial event-related potentials in EEG," *International Journal of Bifurcation and Chaos,* vol. 14, no. 2, pp. 719–726, 2004.

[67] E. Yom-Tov and G. F. Inbar, "Detection of movement related otentials from EEG for possible use in a BCI," *Medical & biological engineering & computing,* vol. 41, pp. 85-93, 2003.

[68] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How Many People are Able to Operate an EEG-Based Brain-Computer Interface (BCI)?," *IEEE transactions on neural systems and rehabilitation engineering,* vol. 11, no. 2, 2003.

[69] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans Biomed Eng,* vol. 51, no. 6, pp. 993-1002, Jun 2004, doi: 10.1109/TBME.2004.827088.

[70] C. Berka and D. J. Levendowski, "Real-Time Analysis of EEG Indexes ofAlertness, Cognition, and Memory Aquired with a Wireless EEG Headset," *Human-Computer Interaction,* vol. 17, no. 2, pp. 151-170, 2004.

[71] S. Sun and C. Zhang, "Adaptive feature extraction for EEG signal classification," *Medical & biological engineering & computing,* Research Support, Non-U.S. Gov't vol. 44, no. 10, pp. 931-5, Oct 2006, doi: 10.1007/s11517-006-0107-4.

[72] M. Dalponte, F. Bovolo, and L. Bruzzone, "Automatic selection of frequency and time intervals for classification of EEG signals," *Electronics Letters,* vol. 43, no. 25, p. 1406, 2007, doi: 10.1049/el:20072428.

[73] S. M. Saddique and L. H. Siddiqui, "EEG Based Brain Computer Interface," *Journal of Software,* vol. 4, no. 6, pp. 550-554, 2009.

[74] N. Brodu, "Exploring two novel features for EEG based Brain computer interfaces, Multifractal cumultants and predictive complexity," *Journal of Neurocomputing,* vol. 79, pp. 87-94, 2010.

[75] B. Awwad Shiekh Hasan and J. Q. Gan, "Unsupervised movement onset detection from EEG recorded during self-paced real hand movement," *Medical & biological engineering & computing,* Research Support, Non-U.S. Gov't vol. 48, no. 3, pp. 245-53, Mar 2010, doi: 10.1007/s11517-009-0550-0.

[76] "Biosemi ActiveTwo." https://www.biosemi.com

[77] K. H. Knuth, M.l Habeck, N. K. Malakar, A. M. Mubeen, and B. Placek, "Bayesian Evidence and Model Selection," *Digital Signal Processing,* vol. 47, 2014.

[78] Y. Wang, Y. T. Wang, and T. P. Jung, "Visual stimulus design for high-rate SSVEP BCI," *Electronics Letters,* vol. 46, no. 15, p. 1057, 2010, doi: 10.1049/el.2010.0923.

[79] G. Bin, "VEP-based brain-computer interfaces: time, frequency, and code modulations," *Journal of Computational Intelligence,* vol. 4, no. 4, pp. 22-26, 2009.

[80] M. Donnerer and A. Steed, "Using a P300 Brain–Computer Interface in an Immersive Virtual Environment," *Presence, MIT,* vol. 19, no. 1, pp. 12-24, 2010.

[81] S. Andrews, R. Palaniappan, A. Teoh, and L. C. Kiong, "Enhancing P300 Component by Spectral Power Ratio Principal Components for a Single Trial Brain-Computer Interface," *American Journal of Applied Sciences,* vol. 5, no. 6, pp. 639-644, 2008.

[82] C. S. Syan and R. Harnarinesingh, "Comparison of Pre-Processing and Classification Techniques for Single-Trial and Multi-Trial P300-Based Brain Computer Interfaces," *American Journal of Applied Sciences* vol. 7, no. 9, pp. 1219-1225, 2010.

[83] C. Groenegress, C. Holzner, C. Guger, and M. Slater. (2010) Effects of P300-Based BCI Use on Reported Presence in a Virtual Environment. *Presence, MIT*.

[84] S. Frenzel and E. Neubert, "Is the P300 Speller Independent?," *Computer Science - Human-Computer Interaction,* vol. arXiv:1006.3688, 2010.

[85] J. Faller and G. Muller-Putz, "An application framework for controlling an avatarin desktop based virtual environment via a software SSVEP brain computer interface," *Presence: Teleoperators and Virtual Environments,* vol. 19, no. 1, pp. 25-34, 2010.

[86] Y. Gu, O. F. do Nascimento, M. F. Lucas, and D. Farina, "Identification of task parameters from movement-related cortical potentials," *Medical & biological engineering & computing,* Research Support, Non-U.S. Gov't vol. 47, no. 12, pp. 1257-64, Dec 2009, doi: 10.1007/s11517-009-0523-3.

[87] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *Journal of neural engineering,* Review vol. 4, no. 2, pp. R32-57, Jun 2007, doi: 10.1088/1741-2560/4/2/R03.

[88] R. Hari, "Action-perception connection and the cortical mu rhythm," *Prog Brain Res,* vol. 159, pp. 253-60, 2006, doi: 10.1016/S0079-6123(06)59017-X.

[89] C. Cooney, Raffaella Folli, and Damien Coyle., "Neurolinguistics research advancing development of a direct-speech brain-computer interface," *iScience,* vol. 8.

[90] A. Porbadnigk, M. Wester, J. P. Calliess, and T. Schultz, "EEG-based speech recognition: Impact of temporal effects," in *Bio-Inspired Systems and Signal Processing*, 2009.

[91] S. Deng, et al., "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," *Journal of neural engineering,* vol. 7, no. 4, 2010.

[92] Y. Song and F. Sepulveda, "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 *Biomedical Circuits and Systems Conference (BioCAS)*, 22-24 Oct. 2014 2014, pp. 568-571, doi: 10.1109/BioCAS.2014.6981789.

[93] S. Zhao, and Frank Rudzicz, "Classifying phonological categories in imagined and articulated speech," presented at the *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[94] S. Iqbal, et al, "Time domain analysis of EEG to classify imagined Speech," presented at the *Proceedings of the Second International Conference on Computer and Communication Technologies*, 2016.

[95] N. Yoshimura, et al., "Decoding of covert vowel articulation using electroencephalography cortical currents," *Frontiers in neuroscience,* vol. 10, 2016.

[96] E. F. González-Castañeda, et al, "Sonification and textification: Proposing methods for classifying unspoken words from EEG signals," *Biomedical Signal Processing and Control,* vol. 37, pp. 82-91, 2017.

[97] A. R. Sereshkeh, et al, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 25, no. 12, pp. 2292-2300, 2017.

[98] N. Hashim, Aziah Ali, and Wan-Noorshahida Mohd-Isa, "Word-based classification of imagined speech using EEG," in *International Conference on Computational Science and Technology*, 2017.

[99] M. Alsaleh, Roger Moore, Heidi Christensen, and Mahnaz Arvaneh, "Examining Temporal Variations in Recognizing Unspoken Words using EEG Signals," presented at the *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018.

[100] A. Rezazadeh Sereshkeh, Rozhin Yousefi, Andrew T. Wong, Frank Rudzicz, and Tom Chau, "Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech," *Brain-Computer Interfaces,* pp. 1-13, 2019.

[101] C. Cooney, Attila Korik, Folli Raffaella, and Damien Coyle, "Classification of imagined spoken word-pairs using convolutional neural networks," presented at the *The 8th Graz BCI Conference*, 2019.

[102] C. Cooney, Folli Raffaella, and Damien Coyle, "Optimizing Input Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG," presented at the *IEEE International Conference on Systems, Man, and Cybernetics*, 2019.

[103] P. Saha, Muhammad Abdul-Mageed, and Sidney Fels, "SPEAK YOUR MIND! Towards Imagined Speech Recognition With Hierarchical Deep Learning," *arXiv:1904.05746,* 2019.

[104] E. Imani, Ali Pourmohammad, Mahsa Bagheri, and Vida Mobasheri, "ICA-Based Imagined Conceptual Words Classification on EEG Signals," *Journal of medical signals and sensors,* vol. 7, no. 3, 2020.

[105] M. A. Bakhshali, Morteza Khademi, Abbas Ebrahimi-Moghadam, and Sahar Moghimi, "EEG signal classification of imagined speech based on

Riemannian distance of correntropy spectral density," *Biomedical Signal Processing and Control,* vol. 59, 2020.

[106] G. Krishna, et al, "State-of-the-art speech recognition using eeg and towards decoding of speech spectrum from eeg," *ArXiv preprint arXiv:1908.05743,* 2019.

[107] P. L. Cornelissen, M. L. Kringelbach, A. W. Ellis, C. Whitney, I. E. Holliday, and P. C. Hansen, "Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography (MEG)," *PloS one,* Research Support, Non-U.S. Gov't vol. 4, no. 4, p. e5359, 2009, doi: 10.1371/journal.pone.0005359.

[108] L. Kauhanen *et al.*, "EEG and MEG brain-computer interface for tetraplegic patients," *IEEE Trans Neural Syst Rehabil Eng,* vol. 14, no. 2, pp. 190-3, Jun 2006, doi: 10.1109/TNSRE.2006.875546.

[109] Y. Bekhti, D. Strohmeier, M. Jas, R. Badeau, and A. Gramfort, "M/EEG source localization with multi-scale time-frequency dictionaries," in *6th International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 2016, pp. 31-35.

[110] D. Gutiérrez, "Using EEG/MEG Data of Cognitive Processes in Brain-Computer Interfaces," in *Medical Physics: Tenth Mexican Symposium on Medical Physics*, 2008, pp. 31-36.

[111] J. Wolpaw and G. Loeb, "BCI Meeting 2005—Workshop on Signals and Recording Methods," *IEEE Trans. Neural Systems and Rehabil. Eng.,* vol. 14, pp. 138-141, 2006.

[112] S. H. Choi and M. Lee, "Estimation of motor imaginary using FMRI experimn based sensor location," *International Journal of Computational Intelligence Research,* vol. 3, no. 1, pp. 46-49, 2007.

[113] G. R. Szycik *et al.*, "Maladaptive connectivity of Broca's area in schizophrenia during audiovisual speech perception: an fMRI study," *Neuroscience,* vol. 253, pp. 274-82, Dec 3 2013, doi: 10.1016/j.neuroscience.2013.08.041.

[114] E. Klein, U. Domahs, M. Grande, and F. Domahs, "Neuro-cognitive foundations of word stress processing - evidence from fMRI," *Behavioral and brain functions : BBF,* Research Support, Non-U.S. Gov't vol. 7, p. 15, May 16 2011, doi: 10.1186/1744-9081-7-15.

[115] S. Coyle, T. Ward, C. Markham, and G. McDarby, "On the suitability of near-infrared (NIR) systems for next-generation brain–computer interfaces," *Physiological Measurement,* vol. 25, pp. 815-822, 2004.

[116] X. Yin and B. Xu, "NIRS-based classification of clench force and speed motor imagery with the use of empirical mode decomposition for BCI.," *Med Eng Phys,* vol. 37, no. 3, pp. 280-286, 2015.

[117] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors,* Research Support, Non-U.S. Gov't
Review vol. 12, no. 2, pp. 1211-79, 2012, doi: 10.3390/s120201211.

[118] B. Graimann, *Brain-Computer Interfaces Revolutionizing Human-Computer Interaction*. New York: Springer, 2010.

[119] R. M. Lakshmi, V. T. Prasad, and C. V. Prakash, "Survey on EEG signal processing methods.," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 4, no. 1, pp. 184-191, 2014.

[120] A. Erfanlan and B. Mahmoudi, "Real-time ocular artifact suppression using recurrent neural network for electro-encephalogram based brain-computer interface," *Med. Biol. Eng. Comput,* vol. 43, pp. 296-305, 2005.

[121] A. Asadi Ghanbari, M. R. Nazari Kousarrizi, M. Teshnehlab, and M. Aliyari, "AN Evolutionary Artifact Rejection Method For Brain Computer Interface Using ICA," *International Journal of Electrical & Computer Sciences IJECS,* vol. 9, no. 9, pp. 461-66, 2011.

[122] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods,* vol. 134, no. 1, pp. 9-21, Mar 15 2004, doi: 10.1016/j.jneumeth.2003.10.009.

[123] G. Gómez-Herrero, *Automatic artifact removal (AAR) toolbox v1. 3 for MATLAB.* 2007.

[124] R. Yang, A. Song, and B. Xu, "Feature extraction of motor imagery EEG based on wavelet transform and higher-order statistics," *International Journal of Wavelets, Multiresolution and Information Processing,* vol. 8, no. 3, pp. 373-384, 2010.

[125] P. M. Lanke, R. K. Shastri, and S. D. Biradar, "EEG signal processing techniques for mental task classification," *International journal of Advanced computing and Electronics Technology,* vol. 2, no. 1, pp. 66-74, 2015.

[126] A. R. Mane, S. D. Biradar, and R. K. Shastri, "Review paper on Feature Extraction Methods for EEG Signal Analysis," *International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS),* vol. 2, no. 1, pp. 545-552, 2015.

[127] N. N. Mbeledogu, M. Odoh, and M. N. Umeh, "Stock Feature Extraction Using Principal Component Analysis," in *International Conference on Computer Technology and Science*, 2012: IACSIT Press, pp. 236-241.

[128] J. Kim, S. Lee, and B. Lee, "Classifying the speech response of the brain using Gaussian hidden markov model (HMM) with independent component analysis (ICA)," in *EMBS*, 2013, pp. 4291-94.

[129] S. Blanco, C. E. D'Attellis, S. I. Isaacson, O. A. Rosso, and R. O. Sirne, "Time-frequency analysis of electroencephalogram series. II. Gabor and wavelet transforms," *Physical Review E,* vol. 54, no. 6, pp. 6661-6672, 1996.

[130] N. Brodu, F. Lotte, and A. Lécuyer, "Comparative study of band-power extraction techniques for motor imagery classification.," in *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, 2011.

[131] M. K. Ahirwal and N. D. londhe, "Power Spectrum Analysis of EEG Signals for Estimating Visual Attention," *International Journal of Computer Applications,* vol. 42, no. 15, 2012.

[132] A. Vuckovic and F. Sepulveda, "Delta band contribution in cue based single trial classification of real and imaginary wrist movements," *Medical & biological engineering & computing,* vol. 46, no. 6, pp. 529-39, Jun 2008, doi: 10.1007/s11517-008-0345-8.

[133] A. Marcano-Cedeño, J. Quintanilla-Domínguez, M. G. Cortina-Januchs, and D. Andina, "Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network," in *IECON - 36th Annual Conference on IEEE Industrial Electronics Society*, Glendale, 2010, pp. 2845-2850.

[134] D. Vidaurre, C. Bielza, and P. Larranaga, "Classification of neural signals from sparse autoregressive features," *Neurocomputing,* vol. 111, pp. 21-26, 2013.

[135] V. Kumar and S. Minz, "Feature Selection: A literature Review," *Smart Computing Review,* vol. 4, no. 3, pp. 211-229, 2014.

[136] K. Sutha and J. J. Tamilselvi, "A Review of Feature Selection Algorithms for Data Mining Techniques," *IJCSE,* vol. 7, no. 6, pp. 63-67, 2015.

[137] J. C. Rojas-Thomas, "New Version of Davies-Bouldin Index for Clustering Validation Based on Cylindrical Distance," in *V Chilean Workshop on Pattern Recognition*, 2013, pp. 81-86.

[138] D. Kan and A. W. Moore, *On Greediness of Feature Selection Algorithms.* . Carnegie Mellon University, The Robotics Institute, 1998.

[139] S. M. Varghese and M. N. Sushmitha, "Efficient Feature Subset Selection Techniques for High Dimensional Data," *IJIRCCE,* vol. 2, no. 3, pp. 3509-15, 2014.

[140] M. Steinbach, Levent Ertöz, and Vipin Kumar, "The challenges of clustering high dimensional data," *New directions in statistical physics . Springer, Berlin,* pp. 273-309, 2004.

[141] H.-P. Kriegel, Peer Kröger, and Arthur Zimek., "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering," *ACM Transactions on Knowledge Discovery from Data (TKDD)* vol. 3, no. 1, 2009.

[142] I. Assent, "Clustering high dimensional data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 2, no. 4, pp. 340-350, 2012.

[143] M. Pavithral, and R. M. S. Parvathi, "A Survey on Clustering High Dimensional Data Techniques," *International Journal of Applied Engineering Research,* vol. 12, no. 11, pp. 2893-2899, 2017.

[144] J. Hämäläinen, Susanne Jauhiainen, and Tommi Kärkkäinen, "Comparison of internal clustering validation indices for prototype-based clustering," *Algorithms,* vol. 10, no. 3, 2017.

[145] A. J. Ferreira, and MáRio AT Figueiredo, "Efficient feature selection filters for high-dimensional data," *Pattern Recognition Letters,* vol. 33, no. 13, pp. 1794-1804, 2012.

[146] X. Zhao, Wei Deng, and Yong Shi, "Feature selection with attributes clustering by maximal information coefficient," *Procedia Computer Science,* vol. 17, pp. 70-79, 2013.

[147] A. Vuckovic and F. Sepulveda, "A two-stage four-class BCI based on imaginary movements of the left and the right wrist," *Med Eng Phys,* vol. 34, no. 7, pp. 964-71, Sep 2012, doi: 10.1016/j.medengphy.2011.11.001.

[148] G. W. Æ. Z. W. Æ. W. C. Æ. J. Zhuang, "Classification of surface EMG signals using optimal wavelet packet method based on Davies-Bouldin criterion," *Med Bio Eng Comput,* vol. 44, pp. 865-872, 2006, doi: 10.1007/s11517-006-0100-y.

[149] F. Sepulveda, M. Meckes, and B. A. Conway, "Cluster separation index suggests usefulness of non-motor EEG channels in detecting wrist movement direction intention," presented at the *IEEE Conference on Cybernetics and Intelligent Systems*, 2004.

[150] F. S. Aleksandra Vuckovic, "Delta band contribution in cue based single trial classification of real and imaginary wrist movements," *Med Biol Eng Comput,* vol. 46, pp. 529–539, 2008, doi: 10.1007/s11517-008-0345-8.

[151] *Statistics and Machine Learning Toolbox User's Guide*. The MathWorks, Inc., 2016.

[152] J. Fan, Yingying Fan, and Yichao Wu, "High-dimensional classification," *High-dimensional data analysis,* pp. 3-37, 2011.

[153] Q. Mai, "A review of discriminant analysis in high dimensions," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 5, no. 3, pp. 190-197, 2013.

[154] T. Górecki, and Maciej Łuczak, "Linear discriminant analysis with a generalization of the Moore–Penrose pseudoinverse," *International Journal of Applied Mathematics and Computer Science,* vol. 23, no. 2, pp. 463-471, 2013.

[155] T. Górecki, and Maciej Łuczak, "Evolutionarily Tuned Generalized Pseudo-Inverse in Linear Discriminant Analysis," *Computing and Informatics,* vol. 35, no. 3, pp. 615-634, 2016.

[156] H. Zou, "Classification with high dimensional features," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 11, no. 1, 2019.

[157] J. Liu, et al., "An efficient pseudo-inverse linear discriminant analysis method for face recognition," in *Proceedings of the International Conferences on Neural Information Processing* . 2006.

[158] S. Varma, and Richard Simon., "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics,* vol. 7, no. 1, 2006.

[159] G. Varoquaux, "Cross-validation failure: small sample sizes lead to large error bars," *Neuroimage,* vol. 180, pp. 68-77, 2018.

[160] C. L. Asplund, J. J. Todd, A. P. Snyder, and R. Marois, "A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention," *Nature Neuroscience,* vol. 13, no. 4, pp. 507-514, 2010.

[161] P. Giacometti, K. L. Perdue, and S. G. Diamond, "Algorithm to find high density EEG scalp coordinates and analysis of their correspondence to structural and functional regions of the brain," *Journal of Neuroscience Methods,* vol. 229, pp. 84-96, 2014.

[162] A. Sepúlveda, M. Linné, J. Llamas Alonso, M. A. Guevara, and M. Hernández González, "Increased prefrontal-parietal eeg gamma band correlation during motor imagery in expert video game players," *Actualidades en Psicología Vol. 28 Núm. 117 2014,* 2014.

[163] M. M. Smith, K. E. Weaver, T. J. Grabowski, R. P. Rao, and F. Darvas, "Non-invasive detection of high gamma band activity during motor imagery," *Frontiers in human neuroscience,* vol. 8, p. 817, 2014.

[164] X. Chi, J. B. Hagedorn, D. Schoonover, and M. D'Zmura, "EEG-Based Discrimination of Imagined Speech Phonemes," *International Journal of Bioelectromagnetism,* vol. 13, no. 4, pp. 201-206, 2011.

[165] E. Kraft, B. Gulyas, and E. Poppel, "Neural Correlates of Thinking," 1 ed.: Springer-Verlag Berlin Heidelberg, 2009, pp. 65-139.

[166] M. Korostenskaja, C. Kapeller, K. H. Lee, C. Guger, J. Baumgartner, and E. M. Castillo, "Characterization of cortical motor function and imagery-related cortical activity: Potential application for prehabilitation," *arXiv preprint arXiv:1708.00525,* 2017.

[167] Y. U. Khan and F. Sepulveda, "Brain–computer interface for single-trial EEG classification for wrist movement imagery using spatial filtering in the gamma band," *IET Signal Processing,* vol. 4, no. 5, p. 510, 2010, doi: 10.1049/iet-spr.2008.0235.

[168] C. Schuster *et al.*, "Best practice for motor imagery: a systematic literature review on motor imagery training elements in five different disciplines," *BMC Medicine,* vol. 9, no. 75, pp. 1741-70, 2011.

[169] E. Curran *et al.*, "Cognitive tasks for driving a brain-computer interfacing system: a pilot study," *IEEE Trans Neural Syst Rehabil Eng,* vol. 12, no. 1, pp. 48-54, Mar 2004, doi: 10.1109/TNSRE.2003.821372.

[170] M. Dyson, "Selecting optimal cognitive tasks for control of a brain computer interface," PhD, CSEE, BCI group, University of Essex, University of Essex, 2010.

[171] T. Geng, J. Q. Gan, M. Dyson, C. S. Tsui, and F. Sepulveda, "A novel design of 4-class BCI using two binary classifiers and parallel mental tasks," *Comput Intell Neurosci,* vol. 2008, p. 437306, 2008, doi: 10.1155/2008/437306.

[172] M. Dyson, F. Sepulveda, and J.Q. Gan, "Localisation of cognitive tasks used in EEG-based BCIs," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology,* vol. 121, no. 9, pp. 1481-1493, 2010.

[173] J. Ruescher, O. Iljina, D. M. Altenmuller, A. Aertsen, A. Schulze-Bonhage, and T. Ball, "Somatotopic mapping of natural upper- and lower-extremity movements and speech production with high gamma electrocorticography," *NeuroImage,* Research Support, Non-U.S. Gov't vol. 81, pp. 164-77, Nov 1 2013, doi: 10.1016/j.neuroimage.2013.04.102.

[174] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Communication,* vol. 52, no. 4, pp. 270-287, 2010, doi: 10.1016/j.specom.2009.08.002.

[175] Y. Song and F. Sepulveda, "A novel onset detection technique for brain–computer interfaces using sound-production related cognitive tasks in simulated-online system," *Neural Eng.,* vol. 14, no. 1, p. 12, 2017.

[176] N. Yoshimura, A. Satsuma, M. Sato, C. S. DaSalla, T. Hanakawa, and Y. Koike, "Usability of EEG cortical currents in classification of vowel speech imagery," in *International Conference on Virtual Rehabilitation*, June 27 - 29, 2011 2011.

[177] L. C. Sarmiento, P. Lorenzana, C. J. Cortes, W. J. Arcos, J. A. Bacca, and A. Tovar, "Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation mode," in *Biosignals and Biorobotics Conference*, 26-28 May 2014 2014, pp. 1-4, doi: 10.1109/BRC.2014.6880997.

[178] S. Rajaram, E. C. Lalor, and B. G. Shinn-Cunningham, "A dynamical system model for neural tracking of speech with EEG," in *Neural Engineering*, 2013, pp. 375-378, doi: 10.1109/NER.2013.6695950.

[179] S. Yoshimoto, Y. Washizawa, T. Tanaka, H. Higashi, and J. Tamura, "Toward multi-command auditory brain computer interfacing using speech stimuli," in *Signal & Information Processing Association Annual Summit and Conference*, 3-6 Dec. 2012 2012, pp. 1-4.

[180] C. Herff and T. Schultz, "Automatic Speech Recognition from Neural Signals: A Focused Review," *Frontiers in neuroscience,* vol. 10, 2016.

[181] S. Sanei and J.A. Chambers, *EEG SIGNAL PROCESSING*. John Wiley & Sons Ltd, 2007.

[182] D. M. DeAlmeida, "Using Brain-Computer Interface to understand and decode neural processes associated with imagined speech," MSc dissertation, Biomedical Engineering, Tecnico Lisboa, 2015.

[183] K. Mohanchandra and S. Saha, "A Communication Paradigm Using Subvocalized Speech: Translating Brain Signals into Speech," *Augment Hum Res,* vol. 1, no. 2, pp. 3-17, 2016.

[184] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Neural Eng,* vol. 7, no. 5, p. 10, 2010.

[185] E. M Mugler *et al.*, "Direct classification of all American English phonemes using signals from functional speech motor cortex," *Neural Eng,* vol. 11, p. 8, 2014.

[186] G. D. Schott, "Penfield's homunculus: a note on cerebral cartography," *Journal of neurology, neurosurgery, and psychiatry,* vol. 56, no. 4, pp. 329-333, 1993.

[187] Alycia Cummingsn, Amebu Seddoh, and Brianna Jallo, "Phonological code retrieval during picture naming: Influence of consonant class," *Brain Research,* vol. 1635, pp. 71-85, 2016.

[188] S. Geva, P. S. Jones, J. T. Crinion, C. J. Price, J. C. Baron, and E. A. Warburton, "The neural correlates of inner speech defined by voxel-based lesion–symptom mapping," *Brain : a journal of neurology,* vol. 134, pp. 3071-3082, 2011.

[189] D. B. Fry, *The Physics of Speech* (Cambridge Textbooks in Linguistics). Cambridge University Press Online Publication, 2012. 1979.

[190] F. Lotte *et al.*, "Electrocorticographic representations of segmental features in continuous speech," *Frontiers in human neuroscience,* vol. 9, 2015.

[191] D. Hsu, M. Hsu, H. L. Grabenstatter, G. A. Worrell, and T. P. Sutula, "Characterization of high frequency oscillations and EEG frequency spectra using the damped-oscillator oscillator detector (DOOD)," *arXiv,* vol. 1309, no. 1086, 2013.

[192] F. Pulvermoller, N. Birbaumer, W. Lutzenberger, and B. Mohr, "High-frequency brain activity: Its possible role in attention, perception and language processing," *Progress in Neurobiology,* vol. 52, pp. 427-445, 1997.

[193] S. N. Baker, G. Curio, and R. N. Lemon, "EEG oscillations at 600 Hz are macroscopic markers for cortical spike bursts," *Physiol,* vol. 550, no. 2, pp. 529-534, 2003.

[194] S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations," *Frontiers in human neuroscience,* vol. 7, p. 138, 2013.

[195] E. M. Whitham *et al.*, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20Hz are contaminated," *Clinical Neurophysiology,* vol. 118, pp. 1877-1888, 2007.

[196] A. Jahangiri and F. Sepulveda, "The contribution of different frequency bands in class separability of covert speech tasks for BCIs," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017: IEEE, pp. 2093-2096.

[197] G. Ruffini *et al.*, "ENOBIO dry electrophysiology electrode; first human trial plus wireless electrode system," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007: IEEE, pp. 6689-6693.

[198] G. Gomez-Herrero *et al.*, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," in *Signal Processing Symposium NORSIG*, 2006, pp. 130-133.

[199] C. Xun, H. Chen, and P. Hu, "Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis," *Journal of Applied Mathematics,* vol. 2014, pp. 1-10, 2014, doi: http://dx.doi.org/10.1155/2014/261347.

[200] B. W. McMenamin, A. J. Shackman, L. L. Greischar, and R. J. Davidson, "Electromyogenic artifacts and electroencephalographic inferences revisited," *NeuroImage,* vol. 54, pp. 4-9, 2011.

[201] B. W. McMenamina *et al.*, "Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG," *NeuroImage,* vol. 49, no. 3, pp. 2416-2432, 2010.

[202] Q. Shie and C. Dapang, "Optimal biorthogonal analysis window function for discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 42, no. 3, pp. 694-697, 1994, doi: 10.1109/78.277869.

[203] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 41, no. 7, pp. 2429-2438, 1993, doi: 10.1109/78.224251.

[204] A. R. Webb, *Statistical Pattern Recognition.* John Wiley and Sons Ltd., 2002.

[205] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans Pattern Anal Mach Intell,* vol. 1, no. 2, pp. 224-7, Feb 1979, doi: 10.1109/TPAMI.1979.4766909.

[206] A. Basirat, M. Sato, J. L. Schwartz, P. Kahane, and J. P. Lachaux, "Parieto-frontal gamma band activity during the perceptual emergence of speech forms," *NeuroImage,* Research Support, Non-U.S. Gov't vol. 42, no. 1, pp. 404-13, Aug 1 2008, doi: 10.1016/j.neuroimage.2008.03.063.

[207] I. BioSemi, "ActiveTwo-Multichannel, DC amplifier, 24-bit resolution, biopotential measurement system with active electrodes," 2001.

[208] C. Vidaurre, R. Scherer, R. Cabeza, A. Schlogl, and G. Pfurtscheller, "Study of discriminant analysis applied to motor imagery bipolar data," *Medical & biological engineering & computing,* vol. 45, no. 1, pp. 61-68, Jan 2007, doi: 10.1007/s11517-006-0122-5.

[209] C. Porcaro, M. T. Medaglia, and A. Krott, "Removing speech artifacts from electroencephalographic recordings during overt picture naming," *NeuroImage,* vol. 105, pp. 171-80, Jan 15 2015, doi: 10.1016/j.neuroimage.2014.10.049.

[210] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage,* vol. 34, pp. 1600-1611, 2007.

[211] *Cortical Functions Rererence*. Hong Kong: Trans Cranial Technologies, 2012.

[212] N. S. Dhanjal, L. Handunnetthi, M. C. Patel, and R. J. S. Wise, "Perceptual Systems Controlling Speech Production," *The Journal of Neuroscience,* vol. 28, no. 40, pp. 9969-9975, 2008.

[213] P. K. McGuire, D. A. Silbersweig, R. M. Murray, A. S. David, R. S. J. Frackowiak, and C. D. Frith, "Functional anatomy of inner speech and auditory verbal imagery," *Psychological Medicine,* vol. 26, no. 1, pp. 29-38, 1996.

[214] A. Jahangiri, J. M. Chau, D. R. Achanccaray, and F. Sepulveda, "Covert Speech vs. Motor Imagery: a comparative study of class separability in identical environments," in *Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE*, 2018: IEEE.

[215] S. Lemm, Benjamin Blankertz, Thorsten Dickhaus, and Klaus-Robert Müller, "Introduction to machine learning for brain imaging," *Neuroimage,* vol. 56, no. 2, pp. 387-399, 2011.

[216] N. O. Schiller, M. Bles, and B. M. Jansma, "Tracking the time course of phonological encoding in speech production: an event-related brain potential study," *Cognitive Brain Research,* vol. 17, pp. 819-831, 2003.

[217] A. Jahangiri and F. Sepulveda, "The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data," *Journal of Medical Systems,* vol. 43, no. 20, 2018, doi: https://doi.org/10.1007/s10916-018-1137-9.

[218] G. Gomez-Herrero, Z. Koldovsky, P. Tichavsky , and K. Egiazarian, "A fast algorithm for blind separation of non-Gaussian and time-correlated signals," in *15th European Signal Processing Conference*, 2007.

[219] P. Tichavsky, E. Doron, A. Yeredor, and J. Nielsen, "A computationally affordable implementation of an asymptotically optimal BSS algorithm for AR sources," in *Signal Processing Conference*, 2006, pp. 1-5.

[220] Gyamfi, K.S., Brusey, J., Hunt, A. and Gaura, E., 2017. Linear classifier design under heteroscedasticity in Linear Discriminant Analysis. Expert Systems with Applications, 79, pp.44-52.

[221] Paliwal, K.K. and Sharma, A., 2012. Improved pseudoinverse linear disicriminant analysis method for dimensionality reduction. International Journal of Pattern Recognition and Artificial Intelligence, 26(01), p.1250002.

[222] Raudys, S. and Duin, R.P., 1998. Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. Pattern recognition letters, 19(5-6), pp.385-392.

# *Publications*

# The contribution of different frequency bands in class separability of covert speech tasks for BCIs

Amir Jahangiri (ajahan@essex.ac.uk), Francisco Sepulveda (fsepulv@essex.ac.uk), BCI research Laboratory, School of Computer Science and Electronic Engineering, University of Essex, UK.

***Abstract*— Several recent studies demonstrate the possibility of using user initiated covert speech mental tasks in brain computer interfaces with varying degrees of success, but details of the best frequency features had not been investigated. In this work, ten volunteers in the age range of 22-70 years participated in the experiment. Eight of them were neurologically healthy, one user was dyslexic, and another was autistic. The four words "back", "forward", "left", and "right" were shortened into "BA", "FO", "LE", and "RY", which are phonetically dissimilar and cognitively relevant directional commands. Participants were asked to covertly speak each as soon as the letters appeared on a screen. Volunteers completed five recording runs. During each run the four words were presented in random succession to avoid sequence bias. The recorded EEG data from the ten users were analysed to discover the best features within a Gabor Transform of the signals, i.e., those yielding the highest word-pair classification accuracy for this specific type of linguistic mental activity. Using this BCI, suitable class separability of covert speech tasks is confirmed for all, including disabled users, with consistently high classification accuracy from 72% to 88% in all cases. Like motor imagery tasks, Alpha and Beta band activity were found to contain 12% and 31% of the most important features respectively. Gamma band activity, which indicates high mental functions, contains 57% of the most important features in this study.**

## I. Introduction

Brain-Computer Interface systems (BCIs) measure a user's brain activity by employing a range of devices and methods, to determine the intent of the user. The measured brain activity may be consciously generated by the user, or may be an unconscious neural response to stimuli. Among the different measurement methods, EEG has many advantages, namely it is non-invasive, relatively cheap, and it provides relatively high time resolution [1]. The most commonly used cognitive task in EEG based BCIs is motor imagery, in which the cortical somatotopic representation of different parts of the body is measured. In theses systems, the user imagines moving a limb in a specific way, to generate a command [2, 3]. Although motor imagery provides a useful way to consciously generate distinguishable brain activity, it requires user training and changes in the imagined movement usually occur in time, which could lead to frequent errors in classification [4]. In addition, some disabled users may experience difficulty with motor tasks even if they are imagery-based [5].

Other cognitive tasks such as mental navigation, covert tone production, solving a multiplication problem, imagining a 3D object, and covert syllable production (speech which is internally generated, but not articulated) have also been shown to generate distinct, task specific EEG patterns for BCI use [5, 6].

Language and cognition are closely related processes, and speech is the most natural and intuitive form of human communication. A BCI system designed to understand commands covertly spoken in the user's mind, is highly desirable. Some researchers have used a combination of covert syllable production and motor imagery [7-9]. In recent years, more attention has been dedicated to BCIs based on linguistic tasks alone [10-18]. However, these studies only focus on the motor imagery of covert articulation, and have ignored preceding linguistic processing stages. most neocortical territories in both hemispheres, as well as many subcortical brain regions, are involved in language [19]. Based on the unique cognitive neuroanatomy of an individual, the spatial and temporal activities may vary [20]. Studies conducted with the use of intra-cranial implants confirm high gamma band activity during covert speech tasks [21-23], which has been previously neglected.

In this study, the shortened name of the four directions, back, forward, left, and right, in the form of "BA", "FO", "LE", and "RY" were chosen for covert speech tasks. Rather than using these phonemes as "non-words", participants are informed of the meaning. These words are cognitively appropriate and different from each other and are suitable for a BCI controlled mouse. Like using diverse limbs in motor imagery, the words are chosen to be phonetically dissimilar, each using a unique combination of language muscles when articulated covertly. In order to demonstrate these differences in a quantitative manner, the properties of each consonant and vowel, such as place of articulation and manner of articulation [24] are presented in Fig. 1. For example, the consonant /b/ is voiced, plosive, and bilabial.

**Figure 1.** Properties of the consonants and vowels in the word classes such as place of articulation and manner of articulation.

| Consonants | /b/ | /f/ | /l/ | /r/ | | Vowels | /a/ | /o/ | /e/ | /ai/ |
|---|---|---|---|---|---|---|---|---|---|---|
| Voiced | 1 | 0 | 1 | 1 | | Lips Spread | 1 | 0 | 1 | 1 |
| Voiceless | 0 | 1 | 0 | 0 | | Lips Round | 0 | 1 | 0 | 0 |
| Plosive | 1 | 0 | 0 | 0 | | Tongue Front | 1 | 0 | 1 | 1 |
| Fricitive | 0 | 1 | 1 | 0 | | Tongue Back | 0 | 1 | 0 | 0 |
| Continuant | 0 | 0 | 0 | 1 | | Jaw Open | 1 | 0 | 0 | 1 |
| Bilabial | 1 | 0 | 0 | 0 | | Jaw Half closed | 0 | 1 | 1 | 0 |
| Labiodental | 0 | 1 | 0 | 0 | | Jaw Closed | 0 | 0 | 0 | 1 |
| Alveolar | 0 | 0 | 0 | 0 | | Diphthongue | 0 | 0 | 0 | 1 |
| Paletal | 0 | 0 | 0 | 1 | | Monophthongue | 1 | 1 | 1 | 0 |

## II. METHODS

### A. Experiment Protocol

In this work, experiments were conducted with ten volunteers in the age range of 22-70. One of the volunteers was dyslexic and another was autistic. All volunteers signed a consent form based on the recommendations of the Ethical Committee of the University of Essex. Users were seated in a comfortable armchair facing a 52-inch screen placed approximately two meters away. A graphical user interface was used to give visual timing cues, and task instruction with the display of a word on the screen, which remains visible for one second. The user was asked to covertly speak the word as soon as it appeared on the screen. Each volunteer completed five recording runs. The users were given a few minutes to rest and move their body between runs. Every run contains 12 trials for every word presented in random order, culminating to 60 trials per word in total. For each word, data from three seconds after the timing cue were used for analysis, and a random idle period of 1-5 seconds occurred before the next cue. Fig. 2 illustrates the imagination protocol used in this experiment.

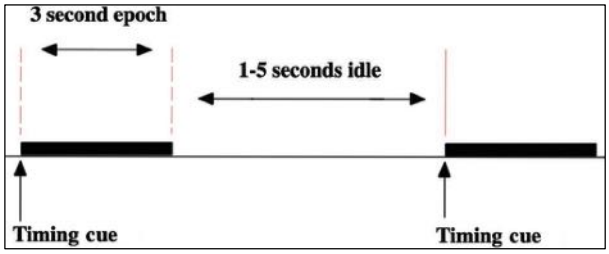
### B. Data Acquisition

The EEG signals were recorded using a 64 channel "Biosemi Active-Two" system [25]. One computer generated the graphical user interface and sent trigger signals to the ActiveTwo device at the instant a time cue was presented to the user. The triggers were sent via the parallel port and were visible in the recorded data. A second computer saved the EEG recordings and was connected to the ActiveTwo's A/D box via USB. Electrode placement was done per the international ABC system, which for 64 channels, this corresponds to the 10/10 system. The ActiveTwo has a pre-amplifier stage on the electrode and can correct for high impedances. However, the offset voltage between the A/D box and the body was kept between 25mV and 50mV as recommended by the manufacturer. Data was recorded at a sampling rate of 2048 Hz to ensure sufficient frequency content, unaffected by the hardware cut-off.

### C. Pre-Processing

The pre-processing was done with the use of the EEGLAB package [26], a MATLAB™ toolbox. At first, the data were down-sampled from 2048 to 256 Hz, and referenced using scalp averages. A band-pass finite impulse response (FIR) filter of order 100, with pass band of (2,100) Hz was then applied. To remove any remaining 50 Hz noise from power lines, an FIR notch filter of order 100, with rejection band of (49,51) Hz was used. EOG and EMG artifacts were removed the Automatic Artifact Rejection (AAR) utility in EEGLAB.

**Figure 2.** Imagination protocol. The user imagines speaking a word when it appears on the screen. Three seconds after each cue are used for analysis, and a random rest period of 1-5 seconds occurs between trials.




In the final stage, epochs were extracted from the EEG data by and for each word, the 60 epochs were saved as a numerical matrix in MATLAB.

### D. Feature Extraction

This work is a novelty search with an exploratory approach. The experiment data are processed offline with no restriction on computational cost. The discrete Gabor transform [27, 28] generates a high-resolution feature space, with no information loss in the time domain, or frequency domain. The definition of the Gabor transform, which is a windowed Fourier transform with a Gaussian window is presented in Fig. 3. For each channel, a 3-second EEG epoch (768 samples), is transformed into a GC matrix with dimensions of 50x96. There are 50 frequency bands of 2 Hz and 96 time steps of 0.03125 seconds.

### E. Feature Selection and classification

In this work, the mean and standard deviation of a ten-fold Block Jack-knife [29] validation method are used to estimate classification accuracy. The data is divided into 10 equal subsets. For each validation fold a different subset is reserved for testing. 9 subsets are used for training the classification object, and the test subset is kept separate and only used to test the performance of the classifier.

Dimensionality reduction of training data and feature selection with clustering algorithms is proven to be extremely effective [30-32]. The Calinski-Harabasz cluster validation criterion was used to determine the potential suitability of each feature in the training data (filter approach) for classification [33]. The most distinct features with best class separation have the largest Calinski-Harabasz indexes, and these values were used to identify an initial subset of the most valuable 1000 features. Definition of the Calinsky-Harabsz index is presented in Fig. 4. The initial subset was further analysed using a cascaded cross validation (wrapper approach) process [34], resulting in a smaller, optimised final subset of best features. The combination of the CHI and CCV methods form a hybrid feature selection process. Using MATLAB™, Pseudo-Linear discriminant analysis was applied for classification of testing data as it consistently out-performed all other supervised machine learning methods for EEG recorded covert speech.

**Figure 3.** Definition of Gabor coefficients by implementation of the direct discrete Gabor transform and a Gaussian window function.

$$GC_{mn} = \sum_{l=0}^{L-1} signal\ (l+1)\ e^{(-2\pi\, lm/M)} conj\ (g(l-an+1))$$

L = length of signal = 768 samples = 3 seconds

a = time step = 8 samples

M = number of frequency channels = 50

m = 0 , ... , 49 frequency index

n = 0 , ... , 95 time index

$g(l) = \sqrt{\frac{\sqrt{2}}{T}} * e^{-\pi (l/T)}$ Gaussian window , $T = \sqrt{a*M}$

**Figure 4.** Definition of the Calinski-Harabasz index. Real and imaginary parts of the features are used to determine the class separation for each feature. More Distinctive features have larger CHIs.

$$CHI = \frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)}$$

$$SS_B = \sum_{i=1}^{k} n_i \left\| m_i - m \right\|^2 \quad , \quad SS_W = \sum_{i=1}^{k} \sum_{x \in x_i} \left\| x - m_i \right\|^2$$

$k$ : clusters = 2 , N : observations

$m_i$ : Centroid of cluster "i"

m : Mean of all data

$\left\| m_i - m \right\|$ : Euclidean distance

## III. Results

### A. Classification Accuracy

There are four words, thus six pairs of word combinations (i.e. "BA" vs. "LE".) The results of pseudo-linear LDA classification true positive rates for the six word combinations are presented in Fig. 5. Classification accuracy for all users in all cases is high (72-88%). The performance of this linguistic BCI is comparable to any system based on motor imagery [7-18]. Participant average performance for all task pairs does not deteriorate; even for users 3, and 5, which are dyslexic and autistic respectively.

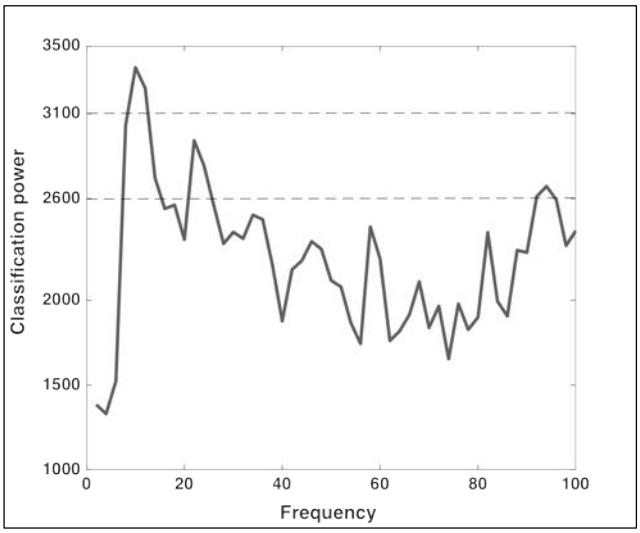
### B. Frequency of best features

The indexes of each Gabor Coefficient, link it to a 2 Hz frequency band. The most important features identified by the hybrid feature selection process, from all users, all word combinations, and all validation folds (a total of 112,353 features) are each placed in their linked band. The sum of features in each band is defined, here, as classification power of that band. Fig. 6 illustrates the classification power of each frequency using all the most important features in this study. The dotted horizontal lines represent one standard deviation, and two standard deviations above mean value. In addition to activity in the Alpha and Beta bands, higher frequencies in the Gamma band are also extremely important. This result confirms recent studies in neurology, linguistics, and psychology, which state that higher mental and cognitive functions such as covert speech require the brain to operate at higher frequencies than motor imagery tasks [35-37].

**Figure 5.** LDA true positive classification accuracy estimated by 10-fold Block Jack-knife validation. Each user completes five recording sessions containing 60 trials for each word. User 3 is dyslexic, and user 5 is autistic.

| | BA-FO | BA-LE | BA-RY | LE-FO | LE-RY | RY-FO |
|---|---|---|---|---|---|---|
| U1 | 80.8±7.9 | 82.4±7.2 | 81.6±5.2 | 86.6±8 | 85±5.2 | 76.7±7.6 |
| U2 | 80±9.7 | 85.8±7.9 | 83.3±7.8 | 82.4±7.2 | 84.1±6.1 | 79.1±8 |
| U3 | 76.6±6.5 | 80±8 | 76.6±7.6 | 83.3±7.8 | 80±4.3 | 83.3±7.8 |
| U4 | 82.5±9.1 | 78.3±4.3 | 75.8±6.1 | 83.3±5.5 | 85±5.2 | 79.1±5.8 |
| U5 | 80.8±4 | 80.8±7.9 | 75.8±8.2 | 78.3±5.8 | 85.8±7.9 | 81.6±12.2 |
| U6 | 80.8±5.6 | 79.1±7 | 78.3±7 | 83.3±7.8 | 80.8±6.8 | 76.6±7.6 |
| U7 | 75.8±4.7 | 80±8.9 | 82.4±8.2 | 77.5±7.9 | 79.1±9 | 82.4±9.1 |
| U8 | 80±8 | 83.3±8.7 | 81.6±8.6 | 75.8±6.1 | 80.8±7.9 | 80±5.8 |
| U9 | 80.8±4 | 72.4±9.6 | 80.8±5.6 | 88.3±7 | 80±7 | 82.4±7.2 |
| U10 | 82.4±10.7 | 82.4±4.7 | 75.8±8.2 | 83.3±7.8 | 81.6±8.6 | 81.6±10.2 |

**Figure 6.** The classification power of each frequency band within the most valuable features. The most important band is Alpha. In addition to alpha and beta band activity, which are also seen in motor imagery, the high gamma band, which indicates higher mental functions, is prominent. The dotted lines show one standard deviation and two standard deviations above mean value.




## IV. Conclusion

The Alpha band has the highest classification power followed by the Beta band. However, only 12% of the most important features (13,894 in total) are in the Alpha, and 31% (34,654 in total) in the Beta band. Although the Gamma band has lower classification power, 57% of the most important features (63,814 in total) are in this band.

Many previous studies on covert speech tasks filter the Gamma band in the pre-processing stage, and eliminate a significant proportion of class dependent information. Thus, the classification accuracy deteriorates. Any future studies in the use of language for BCIs must include the Gamma band.

Within the Gamma band, the classification power increases and peaks at 94 Hz. This may indicate that frequencies above 100 Hz may also contain class dependent activity. However, EEG susceptibility to EOG and EMG artifacts, and limitations of the recording hardware must be carefully considered in future studies if the frequency range of analysis is to be further increased.

## V. Discussion

In this study, the range of analysis in time and frequency domains, and the resolution of Gabor features, represent the absolute minimum requirements of a covert speech BCI. If the frequency range is reduced to 0-90 Hz, the performance of the classifiers deteriorates by 20%. For a range of 0-80 Hz performance falls to level of chance, which is also true if the time range is reduced to 0-2 seconds.

Increasing the resolution of the Gabor transform may potentially improve the performance of the classifier. The Gabor resolution of 2 Hz by 0.03125 seconds provides an acceptable tradeoff between classification performance and available computational power. For any lower Gabor resolution, the performance of the classifier is no better than chance.

## VI. REFERENCES


[1] N. B. Jonathan R. Wolpaw, Dennis J. McFarland, Gert Pfurtscheller, Theresa M. Vaughan, "Brain-computer interfaces for communication and control," *clin Neurophysiology,* vol. 113, pp. 767-791, 2002.

[2] Y. U. Khan and F. Sepulveda, "Brain–computer interface for single-trial EEG classification for wrist movement imagery using spatial filtering in the gamma band," *IET Signal Processing,* vol. 4, p. 510, 2010.

[3] A. Vuckovic and F. Sepulveda, "A two-stage four-class BCI based on imaginary movements of the left and the right wrist," *Med Eng Phys,* vol. 34, pp. 964-71, 2012.

[4] Corina Schuster, Roger Hilfiker, Oliver Amft, Anne Scheidhauer, Brian Andrews, Jenny Butler*, et al.*, "Best practice for motor imagery: a systematic literature review on motor imagery training elements in five different disciplines," *BMC Medicine,* vol. 9, pp. 1741-70, 2011.

[5] E. Curran, P. Sykacek, M. Stokes, S. J. Roberts, W. Penny, I. Johnsrude*, et al.*, "Cognitive tasks for driving a brain-computer interfacing system: a pilot study," *IEEE Trans Neural Syst Rehabil Eng,* vol. 12, pp. 48-54, 2004.

[6] D. Gutiérrez, "Using EEG/MEG Data of Cognitive Processes in Brain-Computer Interfaces," presented at the Medical Physics: Tenth Mexican Symposium on Medical Physics, pp. 31-38, 2008.

[7] Hayrettin Gurkok, Mannes Poel, and J. Zwiers, "Classifying motor imagery in presence of speech," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, Barcelona, pp. 1098-7576, 2010.

[8] Li Wang, Xiong Zhang, and Yu Zhang, "Extending motor imagery by speech imagery for brain-computer interface," presented at the 35th Annual International Conference of the IEEE EMBS, pp. 7056-7059, 2013.

[9] J. Ruescher, O. Iljina, D. M. Altenmuller, A. Aertsen, A. Schulze-Bonhage, and T. Ball, "Somatotopic mapping of natural upper- and lower-extremity movements and speech production with high gamma electrocorticography," *Neuroimage,* vol. 81, pp. 164-77, 2013.

[10] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis and classification of speech imagery EEG for BCI," *Biomedical Signal Processing and Control,* vol. 8, pp. 901-908, 2013.

[11] M. Matsumoto and J. Hori, "Classification of silent speech using adaptive collection," in *Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT), 2013 IEEE Symposium on*, 2013, pp. 5-12, 2013.

[12] M. Matsumoto and J. Hori, "Classification of silent speech using support vector machine and relevance vector machine," *Applied Soft Computing,* vol. 20, pp. 95-102, 2014.

[13] S. Iqbal, Y. U. Khan, and O. Farooq, "EEG based classification of imagined vowel sounds," in *Computing for Sustainable Global Development, 2nd International Conference on*, pp. 1591-1594, 2015.

[14] Katharine Brigham and Vijaya Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication A Preliminary Investigation into Synthetic Telepathy," in *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on* Chengdu, 2010, pp. 1-4, 2010.

[15] A. Riaz, S. Akhtar, S. Iftikhar, A. A. Khan, and A. Salman, "Inter comparison of classification techniques for vowel speech imagery using EEG sensors," in *Systems and Informatics (ICSAI), 2nd International Conference on*, 2014, pp. 712-717, 2014.

[16] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Communication,* vol. 52, pp. 270-287, 2010.

[17] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Netw,* vol. 22, pp. 1334-9, Nov 2009.

[18] Michael D'Zmura, Siyi Deng, Tom Lappas, Samuel Thorpe, and Ramesh Srinivasan, "Toward EEG Sensing of Imagined Speech," in *13th Conference, HCI International* San Diego, USA., 2009, pp. 40-48, 2009.

[19] E. Kraft, B. Gulyas, and E. Poppel, *Neural Correlates of Thinking*. Berlin: Springer-Verlag, 2009.

[20] Peter Indefrey, "The spatial and temporal signatures of word production components: a critical update," *Frontiers in Psychology,* vol. 2, pp. 255-271, 2011.

[21] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," *Front Neurosci,* vol. 5, p. 65, 2011.

[22] A. Llorens, A. Trebuchon, C. Liegeois-Chauvel, and F. X. Alario, "Intra-cranial recordings of brain activity during language production," *Front Psychol,* vol. 2, pp. 1-12, 2011.

[23] Jeremy D. W. Greenlee, Adam W. Jackson, Fangxiang Chen, Charles R. Larson, Hiroyuki Oya, Hiroto Kawasaki*, et al.*, "Human Auditory Cortical Activation during Self- Vocalization," *PLoS One,* vol. 6, pp. 1-15, 2011.

[24] D. B. Fry, *The Physics of Speech*: Cambridge University Press Online Publication, 2012.

[25] *BIOSEMI ActiveTwo*. Available: http://www.biosemi.com/

[26] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods,* vol. 134, pp. 9-21, 2004.

[27] Q. Shie and C. Dapang, "Optimal biorthogonal analysis window function for discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 42, pp. 694-697, 1994.

[28] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 41, pp. 2429-2438, 1993.

[29] A. R. Webb, *Statistical Pattern Recognition.*: John Wiley and Sons Ltd., 2002.

[30] Sherin Mary Varghese and M. N. Sushmitha, "Efficient Feature Subset Selection Techniques for High Dimensional Data," *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 2, pp. 3509-3515, 2014.

[31] K. Sutha and J. Jebamalar Tamilselvi, "A Review of Feature Selection Algorithms for Data Mining Techniques," *International Journal on Computer Science and Engineering,* vol. 7, pp. 63-67, 2015.

[32] Vipin Kumar and Sonajharia Minz, "Feature Selection: A literature Review," *Smart Computing Review,* vol. 4, pp. 211-229, 2014.

[33] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 1650-1654, 2002.

[34] Deng Kan and Andrew W. Moore, *On Greediness of Feature Selection Algorithms.* : Carnegie Mellon University, The Robotics Institute, 1998.

[35] Z. Dan, Enhao Gong, Wei Wu, Jiuluan Lin, Wenjing Zhou, and Bo Hong, "Spoken sentences decoding based on intracranial high gamma response using dynamic time warping," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 3292-329, 2012.

[36] E. C. Leuthardt, X. M. Pei, J. Breshears, C. Gaona, M. Sharma, Z. Freudenberg*, et al.*, "Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task," *Front Hum Neurosci,* vol. 6, pp. 99-102, 2012.

[37] A. Basirat, M. Sato, J. L. Schwartz, P. Kahane, and J. P. Lachaux, "Parieto-frontal gamma band activity during the perceptual emergence of speech forms," *Neuroimage,* vol. 42, pp. 404-13, 2008.

# Covert Speech vs. Motor Imagery: a comparative study of class separability in identical environments

Amir Jahangiri (ajahan@essex.ac.uk), Juan M. Chau (jc18230@essex.ac.uk), David R. Achanccaray (dachanccaray@pucp.pe), Francisco Sepulveda (fsepulv@essex.ac.uk). BCI research Laboratory, School of Computer Science and Electronic Engineering, University of Essex, UK.

***Abstract*— In this study a single experimental protocol and analysis pipeline is used: once for MI tasks, and once for covert speech tasks. The goal of this study is not to maximizing classification accuracy; rather the main objective is to provide an identical environment for both paradigms, while identifying the most important activities related to the most class dependent features. Four volunteers participated in this experiment. With four classes, the average classification accuracy for covert speech tasks is 82.5%, and for motor imagery is 77.2%. The average performance is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect. For motor imagery tasks the most important activities are the execution of imagined movements, and goal driven executive control for suppression of overt movements, which also occur for covert speech tasks. However, the most important activity for covert speech tasks is the linguistic processing stages of word production prior to articulation, which does not occur in motor imagery. These high-Gamma linguistic processes are extremely class dependent, which contribute to the higher performance of covert speech tasks, compared to motor imagery in an otherwise identical environment.**

## I. Introduction

Motor imagery is a well-established paradigm in BCIs. The low-frequency oscillations (< 35 Hz) elicited by "MI" activity, have been detectable by EEG for many decades. MI does not occur independently and is the end-result of many cognitive functions. For example, anticipating an onset cue and initiating "imagined" movement after cue recognition requires stimulus-driven executive control, with high-Gamma activity in regions such as the pre-frontal cortex [1, 2]. To take advantage of such class dependent cognitive activity [3, 4], the entire data bandwidth of the EEG system must be utilized [5] (and not only Alpha and Beta bands). Covert word production begins with high-Gamma (>70 Hz) linguistic processing stages [6-8], followed by motor imagery of articulation [9, 10]. Language is exceedingly more complex than movement [11] and requires analysis with much higher resolution than traditional MI band power [12]. However, covert speech is more intuitive and natural for BCI communication compared to MI. In this study, a single experimental protocol and analysis pipeline is used: once for MI tasks, and once for covert speech tasks. The performance of the system for each paradigm is calculated and the results are discussed.

## II. Methods

### A. Experiment Protocol

In this study, each recording run contains four classes, which are shown in the user interface by four arrows: up, down, left, and right. Within a recording run, 10 examples of each task are presented in a random order (each run has 40 trials) to avoid user fatigue. During recording, a new task is determined by an arrow appearing on the screen for 3 seconds. After the arrow disappears, there is a 3 second standby state. Task onset is presented as a beep sound for all classes. A second beep indicates a rest period before the next trial. The experimental protocol is presented in figure 1.

Each user completes two recording runs, which are identical in every way with the exception of type of mental task (MI, covert speech). For MI tasks, the four arrows represent left hand movement (left arrow), right hand (right arrow), left foot (down arrow), and right foot (up arrow). In covert speech tasks, the user imagines speaking the phonemic structures: BA (back/down arrow), FO (forward/up arrow), LE (left arrow), and RY (right arrow), which are phonetically very dissimilar tasks [13].

### B. Data Acquisition

Four neurologically healthy volunteers participated in this experiment. The EEG signals were recorded using an Enobio dry electrode system with 20 channels and 10/10 configuration [14]. Data was recorded at a sampling rate of 500 Hz and saved in "gdf" format. Compared to wet electrode systems, setting up the Enobio is extremely easy. However, the quality of recorded signals may restrict the number of classes it can use simultaneously. This study provides an evaluation of the system's capability.

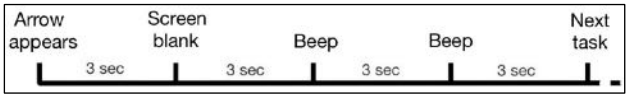



Figure 1. The experiment protocol for recording four randomly presented trials. Each class corresponds to a directional arrow. After task presentation, a beep sound is used for all classes as task onset. A second beep indicates a rest period before the next task.

### C. Pre-Processing

Recorded data was pre-processed using EEGLAB [15]. Data was down-sampled to 256Hz and re-referenced using common average. Line noise was removed with an FIR notch filter (49.5,50.5Hz). The AAR toolkit [16] was used for artefact rejection. EOG and EMG artifacts were reduced, with SOBI [17] and CCA algorithms [18] respectively. These methods outperform ICA, which is ineffective beyond 70 Hz [19, 20]. One-second epochs were extracted from the pre-processed data and saved as a numeric matrix for further analysis.

### D. Feature Extraction

The discrete Gabor transform [21, 22] was used to generate features. The original data can be reconstructed from the features with no information loss. Each Gabor feature contains information on both time and frequency. In this study, a time step of 0.015625 seconds (4 time samples) and a frequency band of 2Hz is used. A one-second epoch from one channel (256 time samples) is converted into a 64x64 feature matrix. Figure 2 presents the definition of the discrete Gabor transform. This method of feature selection makes it possible to identify the type of neural activity from the indexes of the features used in classification.

### E. Feature Selection and classification

Classification true positive rate is estimated by a 5-fold cross validation process [23]. In each fold, 8 trials are used for feature selection and training the classification object, and 2 trials are set aside for testing. The most valuable features for distinguishing four classes are discovered by the Davies-Bouldin index [24]. Initially, all pair-wise DBI matrices are calculated (6 binary combinations with 4 classes). The four-class DBI is a conservative approximation based on the two-class DBIs, which is defined in figure 3. In this experiment, 91% of the total computational cost is spend on generating the DBI matrix. However, the dimensionality of the feature space is significantly reduced. In this study, the 3K most valuable features (from a total of 81920) are identified and used to train the LDA classifier. Features in the test data with the same indexes are used to test the performance of the classifier.

$$GC_{mn} = \sum_{l=0}^{L-1} signal\ (l+1)\ e^{(-2\pi\, lm/M)}\ conj\ (g(l-an+1))$$

L = length of signal = 256 samples = 1 second

a = time step = 4 samples

M = number of frequency channels = 64

m = 0 , ... , 63 frequency index

n = 0 , ... , 63 time index

$$g(l) = \sqrt{\frac{\sqrt{2}}{T}} * e^{-\pi (l/T)}$$ Gaussian window , $T = \sqrt{a*M}$

Figure 2. Definition of Gabor coefficients by implementation of the direct discrete Gabor transform and a Gaussian window function.

| | |
|---|---|
| $DBI = \frac{1}{4}\sum_{i=1}^{4} R_i$ | DB index, four classes |
| $R_i = max_{i \neq j}\ (R_{ij})$ | Conservative approximation |
| $R_{ij} = \frac{S_i + S_j}{D_{ij}}$ | DB index, two classes |
| $D_{ij} = \sqrt{(c_i - c_j)^2}$ | Euclidean distance of centroids |
| $S_i = \frac{1}{N}\sum_{n=1}^{N}(x_n - c_i)$ | Standard deviation of class "i" |

Figure 3. Definition of the Davies-Bouldin index for 4 classes. The most valuable features have the smallest DBI.

## III. Results

### A. Classification Accuracy

The true positive rates for one class vs. all, are estimated as the mean and standard deviation of the five-fold cross validation process. Table 1 contains these results for the four participants and both types of cognitive task. The reader should kindly consider that the objective of this study is not to maximize classification accuracy. The experimental protocol and analysis pipeline provided identical environments for both paradigms, while identifying the most important activities related to the selected features. With four classes, the classification accuracy is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect.

### B. Time-frequency distribution of best features

The 60K features identified in the motor imagery experiments, are shown in a cumulative joint time-frequency plot of the feature space and presented in figure 4. As expected, valuable class dependent activity is not limited to the Alpha and Beta bands. In addition, the nominal bandwidth of (1-125) Hz given by Enobio is confirmed, as valuable features are identified in the entire frequency range.

Table 1. True positive rates of one class vs. all. These are estimated using a five-fold cross validation process. With four classes, the average performance is significantly higher than chance level for both paradigms, suggesting that the results are meaningful, despite being imperfect.

| | *Covert Speech* | *Motor Imagery* |
|---|---|---|
| *User 1* | 85 ± 33.3 | 80.1 ± 32.7 |
| *User 2* | 80.5 ± 30.8 | 68.5 ± 28.1 |
| *User 3* | 87.3 ± 21.2 | 83.4 ± 33.4 |
| *User 4* | 78 ± 18.9 | 78 ± 30.9 |
| *Average* | 82.5 ± 24.1 | 77.2 ± 31.2 |

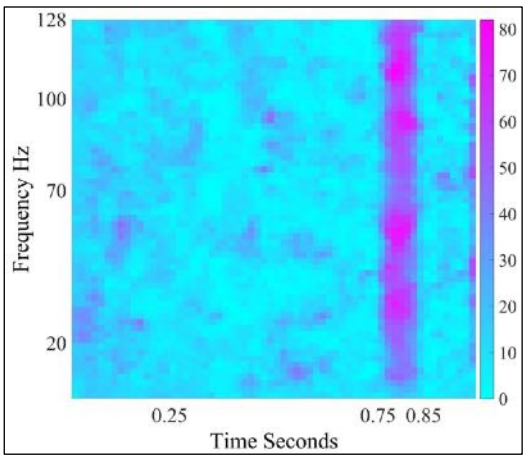



Figure 4. The cumulative joint time-frequency plot of the feature space containing the most valuable 60K features identified in the motor imagery experiments (4 users, five validation folds, and 3K features per fold). The (0.73,0.875) second band contains 15.1% of features.

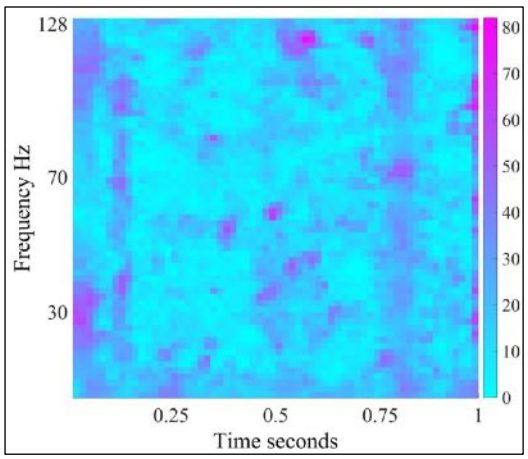



Figure 5. The cumulative joint time-frequency plot of the feature space containing the most valuable 60K features identified in the covert speech experiments (4 users, five validation folds, and 3K features per fold). 48.8% of these features are above 70 Hz. The (0.73,0.875) second band is not as prominent as the MI paradigm from figure 4.

For motor imagery tasks in this experiment 15.1% of all the most valuable features are significantly concentrated within the (0.73-0.875) second range. This time period corresponds with performing imagined movements and the suppression of the Primary Motor Cortex (stopping actual movements) via "goal driven executive control". Such executive control involves high-frequency cognitive activity in brain regions such as the Superior Parietal Cortex and the Pre-Frontal Cortex [1, 2]. 23.2% of all the most valuable features are within the Alpha and Beta bands (MI). The other 76.8% of the features are in the Gamma, and high-Gamma bands (cognitive functions). This suggests that in motor imagery tasks, cognitive functions generate a significantly greater amount of class dependent activity compared to the execution movement.

Figure 5 presents the cumulative joint time-frequency plot of the feature space containing the most valuable 60K features identified in the covert speech experiments. 48.8% of these features are above 70 Hz, which correspond with the linguistic processing functions [8]. These linguistic functions, which are entirely class dependent, do not exist in motor imagery. This provides a possible explanation for the higher classification accuracy of covert speech tasks (82.5%) compared to motor imagery tasks (77.2%) in an identical environment, considering there is a direct positive correlation (with R=0.8822 and P=0) between their performances.

Considering that tasks are identified before trials begin, the cognitively demanding linguistic functions (conceptual preparation, Lemma selection) are completed before onset. The linguistic functions occurring within trials (phonological code retrieval, syllabification) are performed automatically by the brain [9] and require no user effort. All other cognitive functions within trials (executive control, imagined movement) are also present in MI tasks. As a result, the cognitive effort of using covert speech tasks and MI tasks are virtually identical in this study.

## IV. Discussion

The linguistic processing stages of word production prior to articulation, which are entirely class-dependent, consist of conceptual preparation, Lemma selection, phonological code retrieval, and syllabification [25]. By incorporating difference in meaning, and difference in phonetic structure, for selecting selected covert speech tasks, class separability can be significantly enhanced.

In this experiment, linguistic class separability is maximized by selecting phonetically dissimilar covert speech classes [13]. This explains the superior performance of covert speech tasks compared to MI tasks in the otherwise identical environment designed in this study.

Linguistic studies using intra-cranial implants have demonstrated that these linguistic processing stages have high-Gamma signatures in the (70-170Hz) range [10, 26-29]. As bandwidth of EEG systems increases and EMG removal algorithms become more reliable, covert speech BCIs will become much more capable. Although other BCI systems (such as MI) will also improve, language, which is the most intuitive and natural form of human communication, would logically be the preferred paradigm of choice for a BCI.

## V. Acknowledgment

This work was supported by Cienciactiva from Concytec, Peru. Project grant J004-2016 and contract 112-2017.

## VI. References


[1] C. L. Asplund, J. J. Todd, A. P. Snyder, and R. Marois, "A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention," *Nature Neuroscience,* vol. 13, no. 4, pp. 507-514, 2010.

[2] P. Giacometti, K. L. Perdue, and S. G. Diamond, "Algorithm to find high density EEG scalp coordinates and analysis of their correspondence to structural and functional regions of the brain," *Journal of Neuroscience Methods,* vol. 229, pp. 84-96, 2014.

[3] A. Sepúlveda, M. Linné, J. Llamas Alonso, M. A. Guevara, and M. Hernández González, "Increased prefrontal-parietal eeg gamma band correlation during motor imagery in expert video game players," *Actualidades en Psicología Vol. 28 Núm. 117 2014,* 2014.

[4] M. M. Smith, K. E. Weaver, T. J. Grabowski, R. P. Rao, and F. Darvas, "Non-invasive detection of high gamma band activity during motor imagery," *Frontiers in human neuroscience,* vol. 8, p. 817, 2014.

[5] X. Chia, J. B. Hagedorna, D. Schoonovera, and M. D'Zmuraa, "EEG-Based Discrimination of Imagined Speech Phonemes," *International Journal of Bioelectromagnetism,* vol. 13, no. 4, pp. 201-206, 2011.

[6] M. Fukuda, R. Rothermel, C. Juhász, M. Nishida, S. Sood, and E. Asano, "Cortical gamma-oscillations modulated by listening and overt repetition of phonemes," *Neuroimage,* vol. 49, no. 3, pp. 2735–2745, 2010.

[7] V. L.Towle *et al.*, "ECoG gamma activity during a language task: differentiating expressive and receptive speech areas," *Brain,* vol. 131, pp. 2013-2027, 2008.

[8] X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw, and G. Schalk, "Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition," *NeuroImage,* vol. 54, pp. 2960-2972, 2011.

[9] P. Indefrey and W. J. M. Levelt, "The spatial and temporal signatures of word production components," *Cognition,* vol. 92, pp. 101-144, 2004.

[10] E. C. Leuthardt *et al.*, "Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task," *Front Hum Neurosci,* vol. 6, p. 99, 2012.

[11] E. Kraft, B. Gulyas, and E. Poppel, *Neural Correlates of Thinking*. Berlin: Springer-Verlag, 2009, p. 285.

[12] M. Korostenskaja, C. Kapeller, K. H. Lee, C. Guger, J. Baumgartner, and E. M. Castillo, "Characterization of cortical motor function and imagery-related cortical activity: Potential application for prehabilitation," *arXiv preprint arXiv:1708.00525,* 2017.

[13] A. Jahangiri and F. Sepulveda, "The contribution of different frequency bands in class separability of covert speech tasks for BCIs," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 2093-2096: IEEE.

[14] G. Ruffini *et al.*, "ENOBIO dry electrophysiology electrode; first human trial plus wireless electrode system," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 6689-6693: IEEE.

[15] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods,* vol. 134, no. 1, pp. 9-21, Mar 15 2004.

[16] G. Gómez-Herrero, *Automatic artifact removal (AAR) toolbox v1. 3 for MATLAB*. 2007.

[17] G. Gomez-Herrero *et al.*, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," in *Signal Processing Symposium NORSIG*, 2006, pp. 130-133.

[18] X. Chen, C. He, and H. Peng, "Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis," *Journal of Applied Mathematics,* p. 10, 2014.

[19] B. W. McMenamin, A. J. Shackman, L. L. Greischar, and R. J. Davidson, "Electromyogenic artifacts and electroencephalographic inferences revisited," *NeuroImage,* vol. 54, pp. 4-9, 2011.

[20] B. W. McMenamina *et al.*, "Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG," *Neuroimage,* vol. 49, no. 3, pp. 2416-2432, 2010.

[21] Q. Shie and C. Dapang, "Optimal biorthogonal analysis window function for discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 42, no. 3, pp. 694-697, 1994.

[22] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 41, no. 7, pp. 2429-2438, 1993.

[23] A. R. Webb, *Statistical Pattern Recognition.* John Wiley and Sons Ltd., 2002.

[24] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans Pattern Anal Mach Intell,* vol. 1, no. 2, pp. 224-7, Feb 1979.

[25] P. Indefrey, "The spatial and temporal signatures of word production components: a critical update," *Frontiers in Psychology,* vol. 2, no. 255, pp. 255-271, 2011.

[26] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," *Front Neurosci,* vol. 5, p. 65, 2011.

[27] A. Llorens, A. Trebuchon, C. Liegeois-Chauvel, and F. X. Alario, "Intra-cranial recordings of brain activity during language production," *Front Psychol,* vol. 2, p. 375, 2011.

[28] J. D. Greenlee *et al.*, "Human Auditory Cortical Activation during Self- Vocalization," *PLoS One,* vol. 6, no. 3, pp. 1-15, 2011.

[29] A. Basirat, M. Sato, J. L. Schwartz, P. Kahane, and J. P. Lachaux, "Parieto-frontal gamma band activity during the perceptual emergence of speech forms," *Neuroimage,* Research Support, Non-U.S. Gov't vol. 42, no. 1, pp. 404-13, Aug 1 2008.

Journal of Medical Systems (2019) 43:20
https://doi.org/10.1007/s10916-018-1137-9

IMAGE & SIGNAL PROCESSING

# The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data

Amir Jahangiri[1] · Francisco Sepulveda[1]

Received: 23 October 2017 / Accepted: 6 December 2018


**Abstract**
Word production begins with high-Gamma automatic linguistic processing functions followed by speech motor planning and articulation. Phonetic properties are processed in both linguistic and motor stages of word production. Four phonetically dissimilar phonemic structures "BA", "FO", "LE", and "RY" were chosen as covert speech tasks. Ten neurologically healthy volunteers with the age range of 21–33 participated in this experiment. Participants were asked to covertly speak a phonemic structure when they heard an auditory cue. EEG was recorded with 64 electrodes at 2048 samples/s. Initially, one-second trials were used, which contained linguistic and motor imagery activities. The four-class true positive rate was calculated. In the next stage, 312 ms trials were used to exclude covert articulation from analysis. By eliminating the covert articulation stage, the four-class grand average classification accuracy dropped from 96.4% to 94.5%. The most valuable features emerge after Auditory cue recognition (~100 ms post onset), and within the 70–128 Hz frequency range. The most significant identified brain regions were the Prefrontal Cortex (linked to stimulus driven executive control), Wernicke's area (linked to Phonological code retrieval), the right IFG, and Broca's area (linked to syllabification). Alpha and Beta band oscillations associated with motor imagery do not contain enough information to fully reflect the complexity of speech movements. Over 90% of the most class-dependent features were in the 30-128 Hz range, even during the covert articulation stage. As a result, compared to linguistic functions, the contribution of motor imagery of articulation in class separability of covert speech tasks from EEG data is negligible.

**Keywords** Brain-computer interfaces · EEG · Linguistic processing stages · Motor imagery of articulation · Gabor transform · Davies-Bouldin index

## Introduction

Speech is the most natural and intuitive form of human communication. Language and cognition are closely related processes. A BCI system designed to understand commands covertly spoken in the user's mind, is highly desirable. Most neocortical territories in both hemispheres, as well as many subcortical brain regions are involved in language [1]. EEG signals can successfully identify 200–600 Hz cortical spikes [2–4] for medical diagnostic applications. In artefact-free conditions, EEG signals accurately measure induced/evoked high-Gamma brain activity, up to 150 Hz [5–8]. Based on the unique cognitive Neuroanatomy of each individual, the spatial, temporal, and spectral patterns of activity may vary from person to person [9].

Word production begins with semantic (conceptual preparation), lexical (Lemma retrieval), and phonetic (phonological code retrieval and syllabification) linguistic processes, followed by planning the movements of language muscles (phonetic encoding) for articulation [10, 11].

This article is part of the Topical Collection on *Image & Signal Processing*

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s10916-018-1137-9) contains supplementary material, which is available to authorized users.

✉ Amir Jahangiri
ajahan@essex.ac.uk

Francisco Sepulveda
fsepulv@essex.ac.uk

[1] BCI+NE Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

Published online: 18 December 2018

Linguistic phonetic processing is an automatic brain function, which elicits high-Gamma (70–160 Hz) oscillations [12, 13]. In each individual, Phonetic processing activity for a specific word does not change over time [14, 15] and is not affected by priming, cognitive activity, or task frequency [16, 17]. In contrast, semantic and lexical processing, is affected by task frequency, priming, and cognitive activity [18–20], which would also arbitrarily shift the temporal course of all following functions. These problems can be avoided by using a suitable experimental protocol.

In covert speech, the manner of articulation in an individual (their 'accent') is consolidated over time. Covert articulation tasks activate the same language motor centres as their overt form [21, 22]. As a result, covert speech is produced with the same consistency as overt speech. However, in covert speech, the activity of the Primary Motor Cortex is greatly attenuated [23] and may be difficult to detect by EEG. Figure 1, illustrates the functional division of the primary motor cortex, also known as the "Homunculus". Speech production is the most complex motor skill, which takes many years to learn and master. Almost one third of the Primary Motor Cortex is allocated to muscles producing speech, which reflects this complexity [24].

Phonetically dissimilar covert speech tasks create distinctive neural activity associated with the phonological code retrieval and syllabification stages of linguistic processing [25] and involve different language muscle combinations during covert articulation. A linguistic BCI with four classes is sufficiently capable of controlling a smart device with a suitable user interface. In this study, the four directions (back, forward, left, and right) are shortened into Phonemic structures "BA", "FO", "LE", and "RY" and used as covert speech tasks. These covert speech classes are cognitively appropriate directional commands, have little or no overlap with typical mind-wandering states, and provide an intuitive method of communication. For example, the user can move a cursor to the left by covertly speaking "LE". In addition, these Phonemic structures are phonetically dissimilar. To demonstrate these phonetic differences in an accurate and quantitative manner, the properties of each consonant and vowel [26], such as place of articulation and manner of articulation, are presented in Fig. 2 [27]. For example, the consonant /b/ is voiced, plosive, and bilabial.

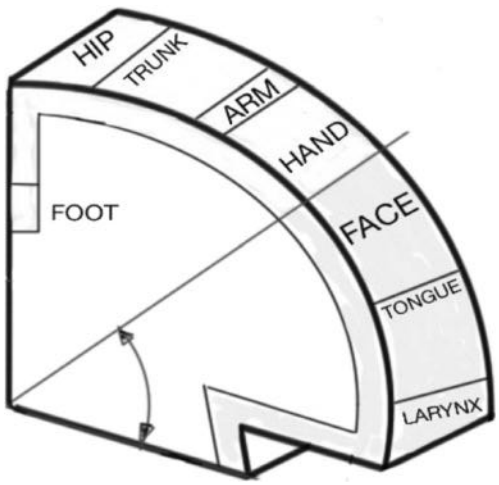



**Fig. 1** The functional division of primary motor cortex. A significant proportion, controls muscles responsible for speech

If the word class is known by the user before the trials, the conceptual preparation stage will be completed in advance. The Lemma selection stage, with multiple competing lemmas will have temporal inconsistencies. If trials are recorded in blocks, only one Lemma is activated and selected. In block recordings, the same auditory time cue, in the form of a "beep' sound, can be used for task onset in all word classes, thus eliminating class-dependent auditory evoked responses from trials. By consolidated the semantic and lexical activities, conceptual preparation and lemma selection are complete before task onset. As a result, trials only contain automatic phonetic linguistic processing stages, and will not be affected by the temporal inconsistency of cognitive activity. Mental effort causes activation of scalp and neck muscles [28], which can mask high-Gamma cortical components. In this work, no mental effort is required from the user during trials. These conditions can be easily reproduced for the online application of this Linguistic BCI, with the same block recordings used for training.

After cue recognition (~100 ms post-onset), the following stages are [23]: Lemma activation (~100-175 ms post-onset), phonological code retrieval (~175-250 ms post-onset) and syllabification (~250-300 ms post-onset). Covert articulation (~500-800 ms post-onset) and the corresponding Motor imagery activity, are separated from the linguistic stages by a ~200 ms interval, during which covert articulation is designed by an internal perceptual process using the working memory and the somatosensory association cortex [9]. Initially one-second trials are used. By using shorter trials (0-312 ms post-onset), the covert articulation stage can be excluded from analysis to study its contribution to classification accuracy.

## Methods

This study was conducted with 10 neurologically healthy volunteers in the age group of 21–33. All volunteers signed a consent form based on the recommendations of the Ethical Committee of the University of Essex. Participants were seated in a comfortable armchair. 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. As a result, Conceptual Preparation, and

| Consonants | /b/ | /f/ | /l/ | /r/ | | Vowels | /a/ | /o/ | /e/ | /ai/ |
|---|---|---|---|---|---|---|---|---|---|---|
| Voiced | 1 | 0 | 1 | 1 | | Lips Spread | 1 | 0 | 1 | 1 |
| Voiceless | 0 | 1 | 0 | 0 | | Lips Round | 0 | 1 | 0 | 0 |
| Plosive | 1 | 0 | 0 | 0 | | Tongue Front | 1 | 0 | 1 | 1 |
| Fricitive | 0 | 1 | 1 | 0 | | Tongue Back | 0 | 1 | 0 | 0 |
| Continuant | 0 | 0 | 0 | 1 | | Jaw Open | 1 | 0 | 0 | 1 |
| Bilabial | 1 | 0 | 0 | 0 | | Jaw Half closed | 0 | 1 | 1 | 0 |
| Labiodental | 0 | 1 | 0 | 0 | | Jaw Closed | 0 | 0 | 0 | 1 |
| Alveolar | 0 | 0 | 0 | 0 | | Diphthongue | 0 | 0 | 0 | 1 |
| Paletal | 0 | 0 | 0 | 1 | | Monophthongue | 1 | 1 | 1 | 0 |

**Fig. 2** Properties of the consonants and vowels in the word classes, such as place of articulation and manner of articulation

Lemma selection are completed before onset. A random rest period between 3 and 7 s was placed between trials to prevent the user from anticipating onset time based on rhythm. This ensures the remaining linguistic activities (Phonological Code Retrieval, Syllabification, Covert Articulation) begin exactly after auditory cue recognition, and the system is perfectly synchronised. Recent studies on the time windows of the processing stages of language production provide evidence of latent activities of over 2000 ms [9]. The 3–7 s idle period sufficiently separates the trials. Figure 3 shows the imagination protocol of the experiment.

The EEG signals were recorded using a 64 channel Biosemi ActiveTwo™ system [29]. One computer generated the graphical user interface and sent trigger signals to the ActiveTwo device at the instant a time cue was presented to the user. The triggers were sent via the parallel port and were visible in the recorded data. A second computer saved the EEG recordings and was connected to the ActiveTwo's A/D box via USB. Electrode placement was done per the international ABC system, which for 64 channels corresponds to the 10/10 system. The ActiveTwo has a pre-amplifier stage on the electrode and can correct for high impedances. However, the offset voltage between the A/D box and the body was kept between 25 mV and 50 mV as recommended by the manufacturer. The data were recorded at a sampling rate of 2048 samples/s, with guaranteed data frequency content of 0-409 Hz according to BioSemi.

The pre-processing was done with the use of EEGLAB [30], an open source MATLAB™ toolbox. Studies conducted with the use of intra-cranial implants confirm high gamma band activity during covert speech tasks [20, 31, 32]. One of the main reasons that numerous studies have failed in achieving high classification accuracy, is that covert speech tasks are treated as motor imagery, and information above the beta band is often ignored or even filtered out [33]. A suitable frequency range (0-128 Hz) for analysing Linguistic activity is achieved by down-sampling the data to 256 Hz. This frequency range is within the operating capability of the ActiveTwo system. The data is then referenced using surface Laplacian. To remove 50 Hz noise from UK power lines, a FIR notch filter, with rejection band of (49.2–50.8 Hz) was applied. Using the Automatic Artifact Removal (AAR) toolbox in EEGLAB [34], EOG and EMG artifacts were reduced, with SOBI [35] and CCA algorithms [36] respectively. These methods

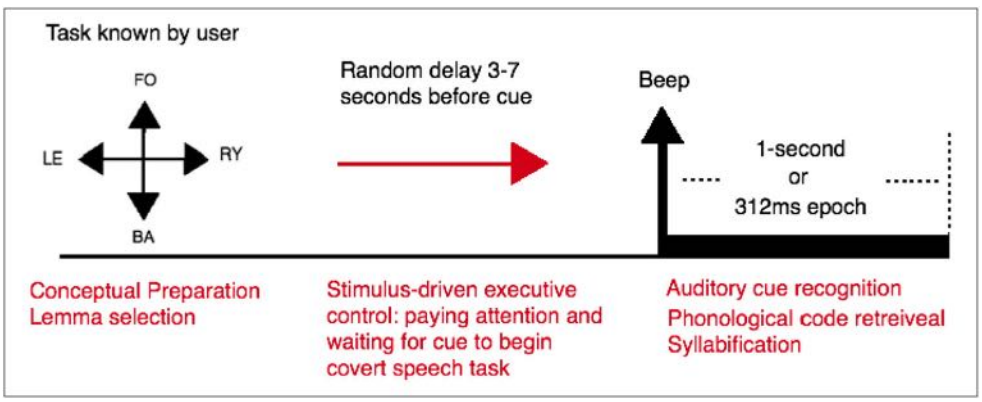



**Fig. 3** Imagination protocol. The user imagines speaking a word when an auditory cue in the form of a beep is presented. One second after each cue are used for the first experiment, and 312 ms for the second. A random rest period of 3–7 s occurs between trials. This sufficiently separated the tasks from one another. Also, the random duration prevents the user from anticipating the task onset based on rhythm. As a result, the next stages of linguistic functions begin exactly after cue recognition and the system is perfectly synchronised

outperform ICA, which is ineffective beyond 70 Hz [37, 38]. Unfortunately, no algorithm can completely eliminate EMG, which elicits 20-200 Hz oscillations in EEG [28, 39]. The most effective solution is to reduce the possibility of recording EMG by controlling the experiment protocol and the environment. The final stage of pre-processing is extracting epochs from the continuous EEG recordings. Each epoch begins when beep sound is generated and ends exactly one second (or 312 ms for shortened trials) later.

This work is a novelty search with an exploratory approach. The experimental data were processed offline and the main objective was to initially create a detailed feature space, in such a way that little or no relevant information is lost or excluded. Features must contain information on time and frequency and should maintain their link to EEG channel for possible topographical analysis. The discrete Gabor Transform [40, 41] (presented in Fig. 4) was thus used as it satisfies all these requirements. Although the Gabor transform is computationally taxing, it has been successfully applied to find hidden information in EEG data with muscle artefact noise contamination to predict onset of seizures [42, 43] and to identify the location of seizure sources [44]. The Gabor transform has also been used for feature generation to classify motor imagery tasks that are very similar, such as different movements of the same hand [45, 46]. In the present study, a time step of 0.03125 s (32 steps per second) and frequency band of 2 Hz (64 frequency bands) were chosen to provide the best tradeoff between classification performance and computational cost.

A 1-s epoch from a single EEG channel (256 samples) is converted into a 64 × 32 feature matrix. For the 312 ms trials (80 samples), one epoch from one channel is converted into a 64 × 10 feature matrix.

To maximise classification accuracy, it is necessary to identify the most distinctive features between the four covert speech classes and use these features to train the classification object. Dimensionality reduction and feature selection with clustering algorithms is proven to be extremely effective [47–49]. The Davies-Bouldin index [50] is a function of within-cluster scatter to between-cluster separation [51, 52], and can be used to determine most useful features to distinguish the four word classes. DBI matrices for all the six word-pairs (e.g., BA vs. FO) are calculated, and used to assign a conservative value to each feature in the "one-vs-all" DBI. Features with the lowest DBI index are considered the most valuable for class separation. Figure 5 shows the definition of the Davies-Bouldin index with four one-dimensional clusters.

The mean and standard deviation of a 10-fold cross validation process [53] were used to estimate the true positive rate. For each validation fold, 27 trials were used for training, and 3 remaining trials were set aside for testing only. Testing trials change from one validation fold to the next, and over 10 folds, all 30 trials are used in testing. The process of cross validation, feature selection, training, and testing used in this work is presented in Fig. 6.

Only the feature generation stage, using the discrete Gabor transform, is applied to the entire dataset. All other calculations are unique and fold-dependent. In this study, for the 1-s trials each DBI matrix has a dimension of 4096 × 32 (64 frequency-bands, 64 channels, 32 time-steps). Based on the DBI, features are ranked and sorted in order of importance. The indexes of the most valuable 4000 features are saved, and these features used for training the classification object. This filtering approach for feature selection reduces the dimensionality of the feature space by 97%, with acceptable computational cost. The 312 ms trials use the same analysis pipeline as 1-s trials. For 64 channels, the dimension of the DBI matrix for 312 ms trials is 4096 × 10 (64 frequency bands, 64 channels, 10 time-steps).

Pseudo-Linear discriminant analysis was applied for classification, as it consistently out-performed all other supervised machine learning methods, for EEG recorded covert speech data [54]. Compared to the training process, the computational cost of testing is negligible.

$$GC_{mn} = \sum_{l=0}^{L-1} signal(l+1)e^{\left(-\frac{2\pi lm}{M}\right)}\text{conj}\left(\text{g}(\text{l}-\text{an}+1)\right)$$

L = length of signal = 256 samples = 1 Second

a = time step of 8 samples

M = number of frequency channels = 64

m = 0 , … , 63 frequency index

n = 0 , … , 31 time index

$$\text{g(l)} = \sqrt{\frac{\sqrt{2}}{T}}e^{-\frac{\pi l}{T}}, \quad T = \sqrt{aM}$$

**Fig. 4** Definition of Gabor coefficients by implementation of the direct discrete Gabor transform and a Gaussian window function

| | |
|---|---|
| $DBI = \frac{1}{4}\sum_{i=1}^{4} R_i$ | DB index, four classes |
| $R_i = max_{i \neq j}\left(R_{ij}\right)$ | Conservative approximation |
| $R_{ij} = \frac{S_i + S_j}{D_{ij}}$ | DB index, two classes |
| $D_{ij} = \sqrt{\left(c_i - c_j\right)^2}$ | Euclidean distance of centroids |
| $S_i = \frac{1}{N}\sum_{n=1}^{N}(x_n - c_i)$ | Standard deviation of class "i" |

**Fig. 5** Definition of the Davies-Bouldin index for 4 one-dimensional clusters. Most valuable features have the smallest DBI

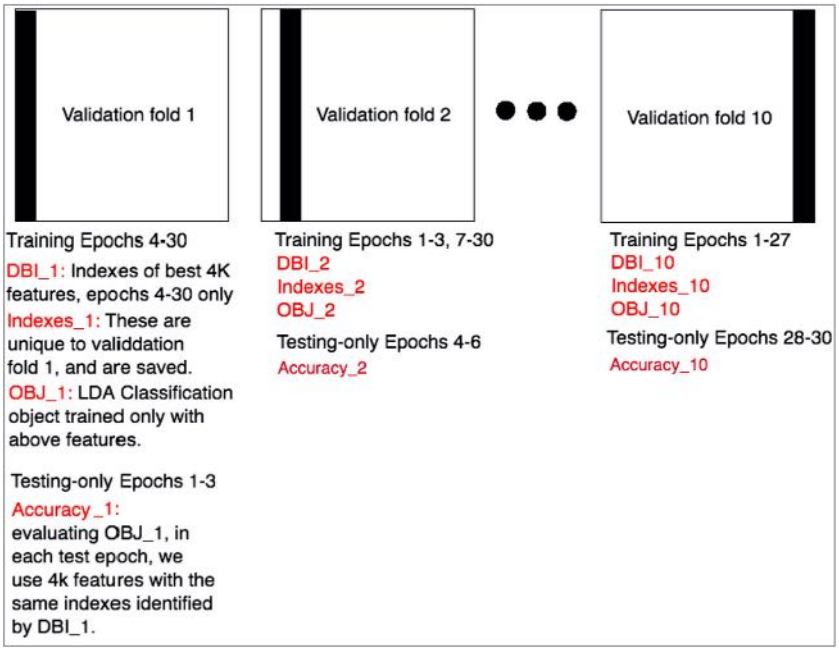



**Fig. 6** The process of cross validation, feature selection, training, and testing used in this work is presented here. The grand average true positive rate is the mean and standard deviation of "Accuracy_1" through "Accuracy_10"

## Results

The true positive rates of one word vs. all, are generated by a standard ten-fold cross validation method. Figure 7 presents these values for 1-s epochs, and for 312 ms epochs. By eliminating the covert articulation stage from trials, the relative contribution of Motor Imagery of speech and linguistic processing stages, in classification accuracy can be determined.

| | 1-second epoch | 312ms epoch |
|---|---|---|
| U1 | 95.8±9% | 96.7±7% |
| U2 | 95±10.5% | 94.1±8.8% |
| U3 | 100% | 100% |
| U4 | 98.3±5.2% | 97.5±5.6% |
| U5 | 94.1±7.9% | 92.5±10% |
| U6 | 93.2±10.9% | 89.1±9.6% |
| U7 | 96.6±8% | 91.6±11.1% |
| U8 | 94.1±7.9% | 89.1±8.8% |
| U9 | 99.1±2.6% | 100% |
| U10 | 97.5±5.6% | 94.1±7.9% |
| Mean | **96.4±2.3%** | **94.5±4%** |

**Fig. 7** The true positive rates of one word vs. all, estimated by a ten-fold cross validation method. Eliminating the covert articulation stage from analysis has less than 2% effect on grand average classification accuracy. Considering the Wilcoxon $p$ value of 0.9269, compared to the high-Gamma linguistic processing stages, the contribution of motor imagery of articulation in class separation of covert speech tasks from EEG data is negligible

The Wilcoxon rank-sum test on both columns returns a p value of 0.9269. By using 312 ms trials instead of 1-s trials to exclude covert articulation, the computational cost is reduced to one third, with less than 2% penalty in classification accuracy. During covert speech, the language motor regions are suppressed, but not completely deactivated [23]. As a result, during the covert articulation stage, there will be minute involuntary muscle movements related to each phonemic structure, which will create class-related, high-Gamma Myoelectric artefacts. The 312 ms trials are complete before the covert articulation stage begins (~500 ms post onset) and are guaranteed to be free from class-related EMG. Possible involuntary early muscle ticks (i.e. lip movements ~160 ms after cue) can cause significant EMG contamination. The CCA algorithm used here, only removes such artefacts from the first 400 ms of data (312 ms trials included) [55].

From 10 users, 10 validation folds/user, and 4000 features/fold, 4e5 best features are identified from the experiment with 1-s trials, and 4e5 from the shortened 312 ms trials. Each Gabor feature is linked to a frequency band, time step, and EEG electrode. The 4e5 features identified in the 1-s trials are cumulatively placed in the $64 \times 32$ feature space to create a colour coded time-frequency representation of the most class-dependent Neural activity, and to identify the electrodes recording this activity for a topographical map of the brain [56, 57]. These plots are illustrated in Fig. 8. The features are highly concentrated in the 70-128 Hz band, even during the covert articulation stage.

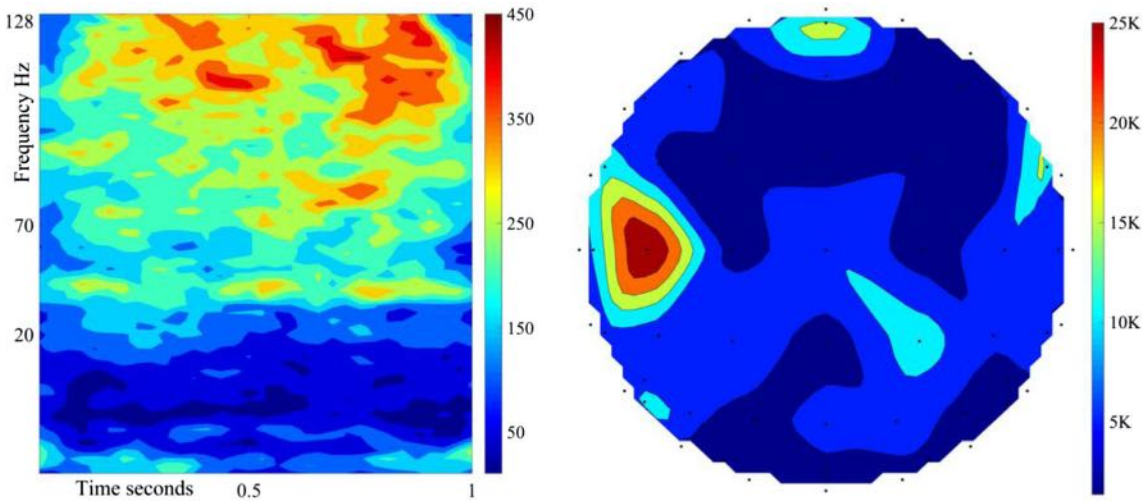



**Fig. 8** The cumulative colour-coded joint time-frequency representation of 4e5 features from 10 users, 1-s trials (Left). The associated topographical plot (Right). The top of the plot is the front of the head. The greatest concentration is within 70–128 Hz

The 4e5 features identified in the 312 ms trials are cumulatively placed in the 64 × 10 feature space to create a colour coded time-frequency representation and used to create a topographic brain map (Fig. 9). The most significant regions are the Prefrontal Cortex [58] (stimulus driven executive control), the left Superior Temporal Gyrus [9] (Wernicke's area, phonological code retrieval), the right, and left Inferior Frontal Gyrus [9] (Broca's area, syllabification). The same regions are prominent in both Figs. 8 and 9.

## Discussion

In a recent publication by these authors [59] an identical experimental protocol and analysis pipeline to this work were used to record mixed randomised trials in a single run using an Enobio dry electrode system with 20 channels. To achieve a manageable recording duration (6–7 min), only 20 trials were recorded per class, and the idle period between trials was reduced to 1–3 s. A grand average classification accuracy of

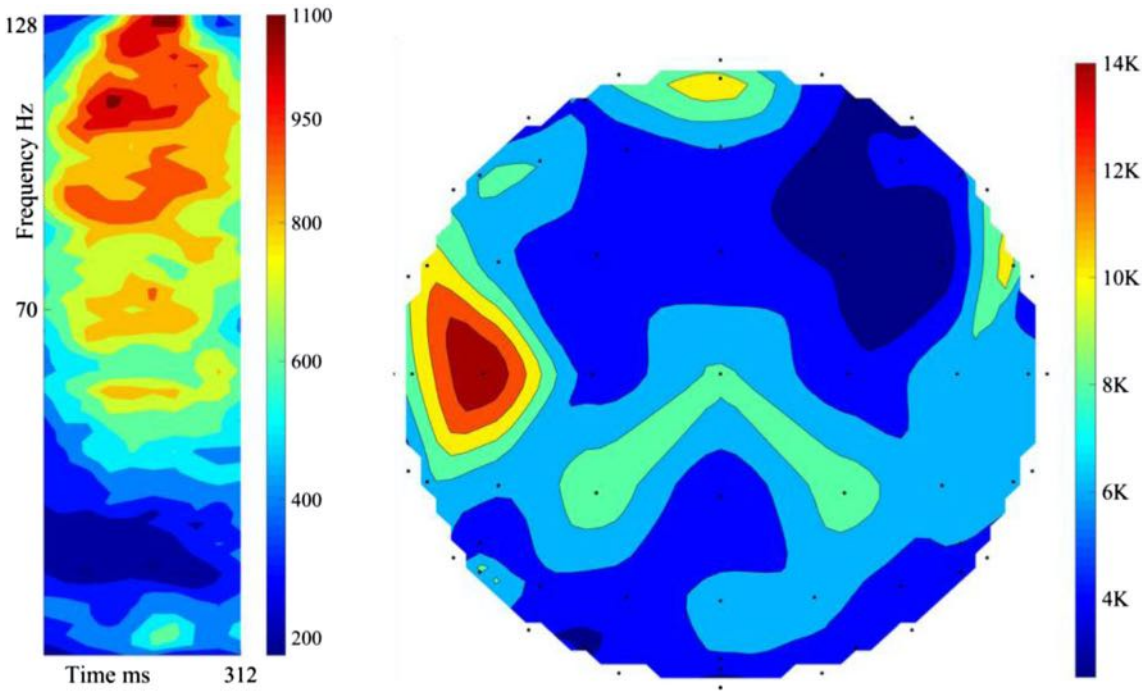



**Fig. 9** The cumulative colour-coded joint time-frequency representation of 4e5 features, 312 ms trials (Left). The associated topographical plot (Right). Most important regions: Prefrontal Cortex, left STG (Wernicke's area), right, and left IFG (Broca's area)

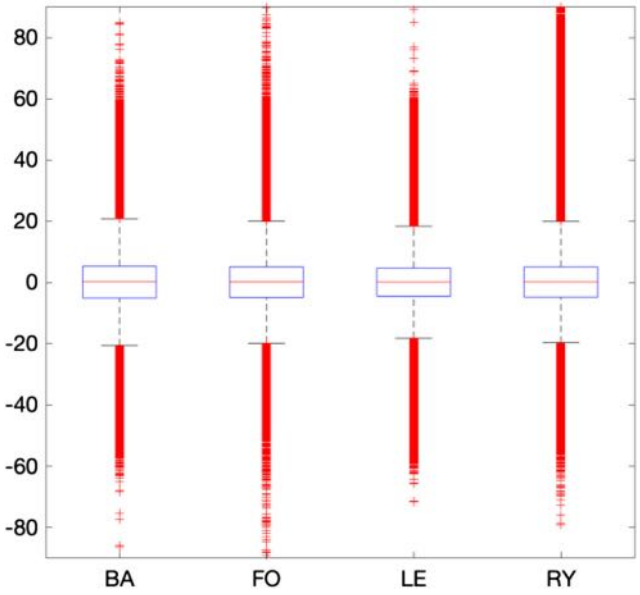



**Fig. 10** The distribution properties of raw EEG recordings in each block for user 1. In all blocks, the mean is 0, std. is 10, the 25% and 75% quartiles are −20 and 20 respectively, and range is near 180. They all have Gaussian distribution. With classification accuracy of 96.7%, no signs of signal drifting exist, suggesting that recording in blocks has little, if any effect on classification accuracy for this data

85% was achieved. Despite using fewer channels, inferior electrodes, and fewer trials compared to the current work, the system performed extremely well for mixed randomised recordings.

Recording 120 trials in a single run using the experimental protocol presented in this work, requires 25–30 min. Maintaining constant focus for such a long duration is exhausting for the user. To reduce user fatigue, trials were recorded in four blocks, each 7–8 min in duration. For each user, the distribution properties (mean, std., rang, etc.) of the raw EEG recordings are virtually identical in all four blocks. Figure 10 presents the distribution properties of the recorded blocks from user 1. The classification accuracy for user 1 is 96.7%. The raw recordings for all 4 blocks closely match each other's distribution properties. This indicates there are no drifts in the recorded signals (i.e. change of an electrode's impedance) causing positive bias in classification accuracy.

The topographical map in Fig. 9 shows the overall activity up to 312 ms post task onset. To demonstrate the sequence of activations, topographical plots with 62 ms intervals are created (Fig. 11). Each plot only contains features from the indicated time range. The sequence of activation is as follows [9]:

- [0-62 ms] Left, and right Auditory Cortex: response to auditory cue.
- [62-124 ms] Prefrontal Cortex [58]: Stimulus-driven executive control, initiating covert speech with auditory cue recognition (100 ms). Left Middle Temporal Gyrus: Lemma activation (100-124 ms).
- [124-186 ms] left Superior Temporal Gyrus: Phonological code retrieval.
- [186-248 ms] Left and right Inferior Frontal Gyrus: syllabification.
- [248-312 ms] Left inferior, and Superior Parietal Cortex [58]: Goal-driven executive control, by suppressing the Primary Motor Cortex, and activating an internal perceptual planning process [60–63].

The syllabification stage is completed sooner than estimated, and the 312 ms trials contain the very early stages of perceptual planning. However, the covert articulation stage, which occurs after the activation of the Supplementary Motor Area [9, 64], is excluded from shortened trials as intended. In the 312 ms trials, the spatial, temporal, and spectral properties of the 4e5 most valuable features identified from 10 participants (Figs. 9 and 11), correspond to the automatic linguistic processing stages of word production prior to articulation, and are supported by a substantial body of evidence [9, 10, 12–15, 20–22, 25, 31, 32, 60]. This, in addition to eliminating the possibility of drifts in the raw EEG recordings, confirm the validity of our findings.

## Conclusions

By excluding motor imagery, grand average classification accuracy dropped from 96.4% to 94.5%. Compared to the high-

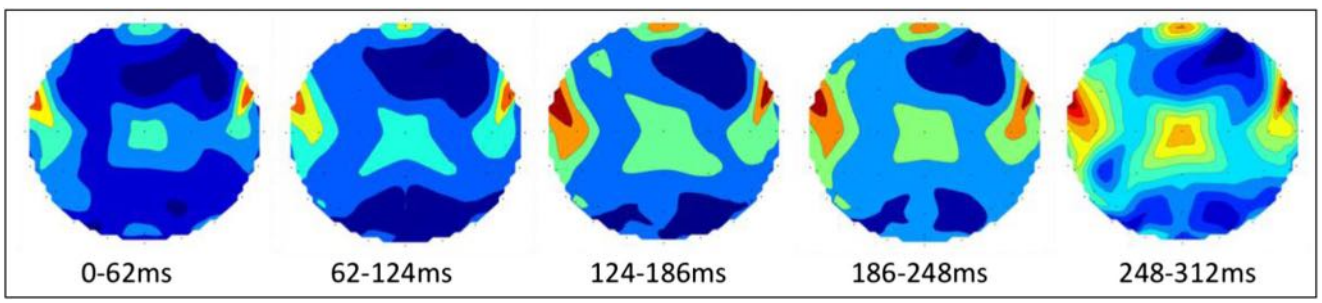



**Fig. 11** Topographical maps of brain regions generating the most distinctive features within the indicated 62 ms interval. The plot for the 248-312 ms interval indicates the early stages of perceptual planning, before activation of the SMA (~500 ms) and covert articulation

Gamma linguistic processing stages of word production, the contribution of motor imagery of articulation in class separability of covert speech tasks is negligible. However, by using 312 ms trials instead of 1-s trials, the computational cost is significantly reduced. The 312 ms trials used in this work, only contain phonetic linguistic processing activity. Phonetic linguistic processing prior to articulation, elicits a unique and word-specific pattern of high-Gamma activity [12, 65], which does not change over time [14, 15] and is not affected by frequency [16] or priming [17]. Phonetic codes are set up and consolidated with the acquisition of language during childhood, and remain unchanged throughout a person's life [17]. Phonetic codes are stored in the long term memory, and are processed automatically by the brain requiring no conscious effort from the user during trials, with immunity from any influence or modification [16, 17, 65, 66]. The experimental protocol and analysis pipeline for 312 ms trials presented in this work can be used as a framework to create an online EEG-based 4-class linguistic BCI in future studies. The raw EEG recordings for all ten participants in this work have been published on "Mendeley Data" (https://doi.org/10.17632/5c2z92vw3g.2) for the benefit of our readers.

## Compliance with ethical standards

**Conflicts of interest** None.

**Ethical approval** Not required.



## References

1. Kraft, E., Gulyas, B., and Poppel, E., Neural correlates of thinking, 1 ed. Berlin, Springer-Verlag, pp. 65–139, 2009.
2. Hsu D., Hsu, M., Grabenstatter, H. L., Worrell, G. A., and Sutula, T. P., Characterization of high frequency oscillations and EEG frequency spectra using the damped-oscillator oscillator detector (DOOD). *arXiv,* vol. 1309, no. 1086, 2013.
3. Pulvermoller, F., Birbaumer, N., Lutzenberger, W., and Mohr, B., High-frequency brain activity: Its possible role in attention, perception and language processing. Prog. Neurobiol. 52:427–445, 1997.
4. Baker, S. N., Curio, G., and Lemon, R. N., EEG oscillations at 600 Hz are macroscopic markers for cortical spike bursts. Physiol 550(2):529–534, 2003.
5. Pitts, M. A., Padwal, J., Fennelly, D., Martínez, A., and Hillyard, S. A., Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. Neuroimage 101:337–350, 2014.
6. Darvas, F., Scherer, R., Ojemann, J. G., Rao, R. P., Miller, K. J., and Sorensen, L. B., High gamma mapping using EEG. Neuroimage 49(1):930–938, 2010.
7. Muthukumaraswamy, S. D., High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Front. Hum. Neurosci. 7:138, 2013.
8. Onton, J., and Makeig, S., High-frequency broadband modulations of electroencephalographic spectra. Front. Hum. Neurosci. 3:61, 2009.
9. Indefrey, P., The spatial and temporal signatures of word production components: a critical update. Front. Psychol. 2(255):255–271, 2011.
10. Indefrey, P., and Levelt, W. J. M., The spatial and temporal signatures of word production components. Cognition 92:101–144, 2004.
11. Leuthardt, E. C. et al., Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. Front. Hum. Neurosci. 6:99, 2012.
12. Fukuda, M., Rothermel, R., Juhász, C., Nishida, M., Sood, S., and Asano, E., Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. Neuroimage 49(3):2735–2745, 2010.
13. Towle, V. L. et al., ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain 131: 2013–2027, 2008.
14. Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., and Knight, R. T., Single-trial speech suppression of auditory cortex activity in humans. J. Neurosci. 30(49):16643–16650, 2010.
15. Creutzfeldt, O., Ojemann, G., and Lettich, E., Neuronal activity in the human lateral temporal lobe. Exp. Brain Res. 77:451–475, 1989.
16. Martin, R. C., Lesch, M. F., and Bartha, M. C., Independence of Input and Output Phonology in Word Processing and Short-Term Memory. J. Mem. Lang. 41:3–29, 1999.
17. Leinenger, M., Phonological coding during reading. Psychol. Bull. 140(6):1534–1555, 2014.
18. Kaan, E., Event related potentials and language processing: a brief overview. Lang Ling Compass 1(6):571–579, 2007.
19. Ganushchak, L. Y., Christoffels, I. K., and Schiller, N. O., The use of electroencephalography in language production research: a review. Front. Psychol. 2:208, 2011.
20. Llorens, A., Trebuchon, A., Liegeois-Chauvel, C., and Alario, F. X., Intra-cranial recordings of brain activity during language production. Front. Psychol. 2:375, 2011.
21. Numminena, J., and Curio, G., Differential effects of overt, covert and replayed speech on vowel- evoked responses of the human auditory cortex. Neurosci. Lett. 272:29–32, 1999.
22. Chakrabarti, S., Sandberg, H. M., Brumberg, J. S., and Krusienski, D. J., Progress in Speech Decoding from the Electrocorticogram. Biomed. Eng. Lett. 5:10–21, 2015.
23. Pei, X., Leuthardt, E. C., Gaona, C. M., Brunner, P., Wolpaw, J. R., and Schalk, G., Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. NeuroImage 54:2960–2972, 2011.
24. Schott, G. D., Penfield's homunculus: a note on cerebral cartography. J. Neurol. Neurosurg. Psychiatry 56(4):329–333, 1993.
25. Cummingsn, A., Seddoh, A., and Jallo, B., Phonological code retrieval during picture naming: Influence of consonant class. Brain Res. 1635:71–85, 2016.
26. Fry, D. B., *The Physics of Speech* (Cambridge Textbooks in Linguistics). Cambridge University Press Online Publication, 2012.
27. Jahangiri, A. and Sepulveda, F., The contribution of different frequency bands in class separability of covert speech tasks for BCIs. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pp. 2093–2096: IEEE, 2017.
28. Whitham, E. M. et al., Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20Hz are contaminated. Clin. Neurophysiol. 118:1877–1888, 2007.

29. I. BioSemi, ActiveTwo-Multichannel, DC amplifier, 24-bit resolution, biopotential measurement system with active electrodes, 2001.
30. Delorme, A., and Makeig, S., EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134(1):9–21, 2004.
31. Brumberg, J. S., Wright, E. J., Andreasen, D. S., Guenther, F. H., and Kennedy, P. R., Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Front. Neurosci. 5:65, 2011.
32. Greenlee, J. D. et al., Human Auditory Cortical Activation during Self- Vocalization. PLoS One 6(3):1–15, 2011.
33. Chi, X., Hagedorn, J. B., Schoonover, D., and D'Zmura, M., EEG-Based Discrimination of Imagined Speech Phonemes. International Journal of Bioelectromagnetism 13(4):201–206, 2011.
34. Gómez-Herrero, G., *Automatic artifact removal (AAR) toolbox v1. 3 for MATLAB.* 2007.
35. Gomez-Herrero, G., *et al.*, Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel. In: *Signal Processing Symposium NORSIG*, pp. 130–133, 2006.
36. Xun, C., Chen, H., and Hu, P., Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis. J. Appl. Math. 2014:1–10, 2014.
37. McMenamin, B. W., Shackman, A. J., Greischar, L. L., and Davidson, R. J., Electromyogenic artifacts and electroencephalographic inferences revisited. NeuroImage 54:4–9, 2011.
38. McMenamina, B. W. et al., Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG. Neuroimage 49(3):2416–2432, 2010.
39. Goncharova, I. I., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R., EMG contamination of EEG: spectral and topographical characteristics. Clin. Neurophysiol. 114:1580–1593, 2003.
40. Shie, Q., and Dapang, C., Optimal biorthogonal analysis window function for discrete Gabor transform. IEEE Trans. Signal Process. 42(3):694–697, 1994.
41. Qian, S., and Chen, D., Discrete Gabor transform. IEEE Trans. Signal Process. 41(7):2429–2438, 1993.
42. Quiroga, R. Q., Blanco, S., Rosso, O. A., Garcia, H., and Rabinowicz, A., Searching for hidden information with Gabor Transform in generalized tonic-clonic seizures. Electroencephalogr. Clin. Neurophysiol. 103(4):434–439, 1997.
43. Blanco, S., D'Attellis, C. E., Isaacson, S. I., Rosso, O. A., and Sirne, R. O., Time-frequency analysis of electroencephalogram series. II. Gabor and wavelet transforms. Phys. Rev. E 54(6):6661–6672, 1996.
44. Bekhti, Y., Strohmeier, D., Jas, M., Badeau, R., and Gramfort, A., M/EEG source localization with multi-scale time-frequency dictionaries. In *6th International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pp. 31–35, 2016.
45. Vuckovic, A., and Sepulveda, F., Delta band contribution in cue based single trial classification of real and imaginary wrist movements. Med. Biol. Eng. Comput. 46(6):529–539, 2008.
46. Vuckovic, A., and Sepulveda, F., Quantification and visualisation of differences between two motor tasks based on energy density maps for brain-computer interface applications. Clin. Neurophysiol. 119(2):446–458, 2008.
47. Varghese, S. M., and Sushmitha, M. N., Efficient Feature Subset Selection Techniques for High Dimensional Data. IJIRCCE 2(3): 3509–3515, 2014.
48. Sutha, K., and Tamilselvi, J. J., A Review of Feature Selection Algorithms for Data Mining Techniques. IJCSE 7(6):63–67, 2015.
49. Kumar, V., and Minz, S., Feature Selection: A literature Review. Smart Computing Review 4(3):211–229, 2014.
50. Davies, D. L., and Bouldin, D. W., A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1(2):224–227, 1979.
51. Rojas-Thomas, J. C., New version of Davies-Bouldin index for clustering validation based on cylindrical distance. In *V Chilean Workshop on Pattern Recognition*, pp. 81–86, 2013.
52. Maulik, U., and Bandyopadhyay, S., Performance evaluation of some clustering algorithms and validity indices. IEEE Trans. Pattern Anal. Mach. Intell. 24(12):1650–1654, 2002.
53. Webb, A. R., Statistical pattern recognition. Hoboken: John Wiley and Sons Ltd., 2002.
54. Vidaurre, C., Scherer, R., Cabeza, R., Schlogl, A., and Pfurtscheller, G., Study of discriminant analysis applied to motor imagery bipolar data. Med. Biol. Eng. Comput. 45(1):61–68, 2007.
55. Porcaro, C., Medaglia, M. T., and Krott, A., Removing speech artifacts from electroencephalographic recordings during overt picture naming. Neuroimage 105:171–180, 15 2015.
56. Jurcak, V., Tsuzuki, D., and Dan, I., 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. NeuroImage 34:1600–1611, 2007.
57. *Cortical Functions Rererence*. Trans Cranial Technologies ldt. Wanchai, Hong Kong, 2012.
58. Asplund, C. L., Todd, J. J., Snyder, A. P., and Marois, R., A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. Nat. Neurosci. 13(4):507–514, 2010.
59. Jahangiri, A., Chau, J. M., Achanccaray, D. R., and Sepulveda, F., Covert speech vs. motor imagery: a comparative study of class separability in identical environments. In: *Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE*, IEEE, 2018.
60. Dhanjal, N. S., Handunnetthi, L., Patel, M. C., and Wise, R. J. S., Perceptual Systems Controlling Speech Production. J. Neurosci. 28(40):9969–9975, 2008.
61. Watkins, K., and Paus, T., Modulation of Motor Excitability during Speech Perception: The Role of Broca's Area. Cogn. Neurosci. 16(6):978–987, 2004.
62. Tian, X., and Poeppel, D., Mental imagery of speech: linking motor and pereptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience.* 2, 2012.
63. Tian, X., and Poeppel, D., Mental imagery of speech and movement implicates the dynamics of internal forward models. Front. Psychol. 1:166, 2010.
64. McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., and Frith, C. D., Functional anatomy of inner speech and auditory verbal imagery. Psychol. Med. 26(1):29–38, 1996.
65. Schiller, N. O., Bles, M., and Jansma, B. M., Tracking the time course of phonological encoding in speech production: an event-related brain potential study. Cogn. Brain Res. 17:819–831, 2003.
66. Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J. C., and Warburton, E. A., The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. Brain 134:3071–3082, 2011.

# A Novel EEG-Based Four-Class Linguistic BCI*

Amir Jahangiri, David Achanccaray, Francisco Sepulveda

*Abstract*— **In this work, we present a novel EEG-based Linguistic BCI, which uses the four phonemic structures "BA", "FO", "LE", and "RY" as covert speech task classes. Six neurologically healthy volunteers with the age range of 19-37 participated in this experiment. Participants were asked to covertly speak a phonemic structure when they heard an auditory cue. EEG was recorded with 64 electrodes at 2048 samples/s. The duration of each trial is 312ms starting with the cue. The BCI was trained using a mixed randomized recording run containing 15 trials per class. The BCI is tested by playing a simple game of "Wack a mole" containing 5 trials per class presented in random order. The average classification accuracy for the 6 users is 82.5%. The most valuable features emerge after Auditory cue recognition (~100ms post onset), and within the 70-128 Hz frequency range. The most significant identified brain regions were the Prefrontal Cortex (linked to stimulus driven executive control), Wernicke's area (linked to Phonological code retrieval), the right IFG, and Broca's area (linked to syllabification). In this work, we have only scratched the surface of using Linguistic tasks for BCIs and the potential for creating much more capable systems in the future using this approach exists.**

## I. Introduction

Language is the most natural and intuitive form of human communication. It is also the very mechanism responsible for cognition and thinking. A BCI system designed to function by understanding the words a user covertly speaks (generated, but not articulated) is highly desirable. However, compared to all other Neural functions used in previous BCIs, Language is the most complex, involving most neocortical territories in both hemispheres, as well as many subcortical brain regions [1]. Word production begins with semantic (conceptual preparation), lexical (Lemma retrieval), and phonetic (phonological code retrieval and syllabification) linguistic processes, followed by planning the movements of language muscles (phonetic encoding) for articulation [2, 3]. The duration of Semantic and Lexical functions is affected by priming, cognitive activity, or task frequency [4-6]. Such temporal inconsistencies can arbitrarily shift the course of all following Linguistic functions. This problem has been solved by using a suitable experimental protocol previously developed and proven by these authors [7-9].

*This work was supported by Fondecyt from Concytec, Peru. Contract 112-2017 and Project grant J004-2016.

Amir Jahangiri (ajahan@essex.ac.uk) and Francisco Sepulveda (fsepulv@essex.ac.uk) are with the BCI+NE Laboratory, School of Computer Science and Electronic Engineering, University of Essex, UK.

David Achanccaray (david.ad@dc.tohoku.ac.jp) is with the Neuro-Robotics Laboratory, Department of Robotics, Tohoku University, Sendai, Japan.

Linguistic Phonetic processing is an automatic brain function, which is highly consistent for each individual and does not change over time [10, 11] and elicits high-Gamma (70-160 Hz) oscillations [12, 13]. This frequency range can be accurately measured by many modern EEG systems [14-17], including the Biosemi ActiveTwo™ system [18] used in this work. The Covert Articulation stage is not analyzed in this experiment, as it has negligible effect on class separability compared to the previous linguistic functions and is highly susceptible to EMG contamination [8]. The novel Linguistic BCI presented here is used by six participants to play a simple 4-class game. The BCI performance is presented and results are discussed.

## II. Methods

### A. Experiment Protocol

This study was conducted with the help of six neurologically healthy volunteers in the age group of 21-33. All volunteers signed a consent form based on the recommendations of the Ethical Committee of the University of Essex. Participants were seated in a comfortable armchair in front of a screen during training the BCI and playing a game of "whack a mole". In this game, an image appears on the screen for one second, in which a mole pops its head out of one four holes in the ground (left, right, back, forward). The user is informed of the task before each run. As a result, Conceptual Preparation, and Lemma selection are completed before onset. The user then waits for an auditory cue to covertly speak a command and move a hammer to the correct location to whack the mole. The four chosen covert speech tasks are "BA", "FO", "LE", and "RY", which are Phonetically very dissimilar [7-9] and can be used as intuitive shortened directional commands. For example, the hammer can hit the left hole by using the covert speech task "LE". For all classes, an identical "Beep" sound was used as the auditory cue. After task presentation, a random rest period between 1 and 3 seconds was placed before the auditory cue to prevent the user from anticipating onset time based on rhythm. This ensures the following linguistic activities begin exactly after auditory cue recognition (trigger driven executive control), and the system is perfectly synchronized. By using 312ms epochs, the only analyzed Linguistic functions are Phonological Code Retrieval and Syllabification, which are performed automatically by the brain and have an unchangeable duration unique to each individual participant. The experiment has two stages: training and testing. During the training stage, 15 trials of each of the four classes (60 trials in total) are presented to the user in a mixed randomised order. For the testing stage, 5 trials of each of the four classes (20 trials in total) are presented to the user in a mixed randomised order. Figure 1 illustrates the experimental protocol.

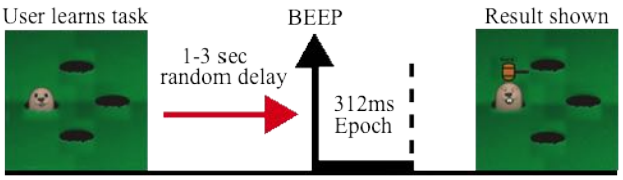



***Figure 1.*** *Imagination protocol. An image presented for 1 second shows the mole in one of the holes (left in this case). The user chooses the correct direction for landing the hammer and the correct covert speech task to perform (in this case "LE"). After the task image disappears the user waits to hear the beep sound to begin covertly speaking the correct word. In the testing stage the result is also shown in an image, completing the cycle.*

### B. Data Acquisition

The EEG signals were recorded using a 64 channel Biosemi ActiveTwo™ system [18]. A computer generated the graphical user interface and sent trigger signals to the ActiveTwo device at the instant a time cue was presented to the user. The triggers were sent via the parallel port and were visible in the recorded data. In addition, special triggers start and stop recording, which enable the computer generating the GUI to also save and analyze the EEG recordings via USB connection to the ActiveTwo's A/D box. Electrode placement was done per the international ABC system, which for 64 channels corresponds to the 10/10 system. Two facial electrodes were used to detect ocular or other muscle artefacts. The ActiveTwo has a pre-amplifier stage on the electrode and can correct for high impedances. However, the offset voltage between the A/D box and the body was kept between 25mV and 50mV as recommended by the manufacturer. The data were recorded at a sampling rate of 2048 samples/s, with guaranteed data frequency content of 0-409Hz according to BioSemi.

### C. Pre-Processing

The pre-processing was done with the use of EEGLAB [19], an open source MATLAB™ toolbox. Studies conducted with the use of intra-cranial implants confirm high gamma band activity during covert speech tasks [6, 20, 21]. One of the main reasons that numerous studies have failed in achieving high classification accuracy, is that covert speech tasks are treated as motor imagery, and information above the beta band is often ignored or even filtered out [22]. A suitable frequency range (0-128Hz) for analysing Linguistic activity is achieved by down-sampling the data to 256Hz. This frequency range is within the operating capability of the ActiveTwo system. The data is then referenced using surface Laplacian. To remove 50Hz noise from UK power lines, a FIR notch filter, with rejection band of (49.2-50.8 Hz) was applied. Using the Automatic Artifact Removal (AAR) toolbox in EEGLAB [23], EOG and EMG artifacts were reduced, with SOBI [24] and CCA algorithms [25] respectively. Finally, epochs are extracted beginning with the beep sound, ending 312ms later.

### D. Feature Extraction and Selection

The discrete Gabor Transform [26, 27] was used to generate features as Gabor coefficients contain information on time and frequency and maintain their link to their source EEG channel for possible topographical analysis. Equation 1 presents the definition of the Gabor transform.

$$GC_{mn} = \sum_{l=0}^{L-1} signal(l+1)e^{\left(-\frac{2\pi lm}{M}\right)}\text{conj}\left(\text{g}(\text{l}-\text{an}+1)\right)$$

L = length of signal

a = time step of 8 samples

M = number of frequency channels = 64

m = 0 , … , 63 frequency index

n = 0 , … , 9 time index

$$\text{g(l)} = \sqrt{\frac{\sqrt{2}}{T}}e^{-\frac{\pi l}{T}}, \quad T = \sqrt{aM}$$

***Equation 1.*** *Definition of Gabor coefficients by implementation of the direct discrete Gabor transform and a Gaussian window function.*

In the present work, a time step of 0.03125 seconds (32 steps per second) and frequency band of 2Hz (64 frequency bands) were chosen to provide the best tradeoff between classification performance and computational cost [7-9]. A 312ms epoch from a single EEG channel (80 samples) is converted into a 64x10 feature matrix. The Davies-Bouldin index [28] is a function of within-cluster scatter to between-cluster separation [29, 30], and can be used to determine most useful features to distinguish the four word classes. DBI matrices for all the six word-pairs (e.g., BA vs. FO) are calculated, and used to assign a conservative value to each feature in the "one-vs-all" DBI. Features with the lowest DBI index are considered the most valuable for class separation.

In this study, 312ms trials are converted into a DBI matrix with a dimension of 4096x10 (64 frequency-bands, 64 EEG channels, 10 time-steps). Equation 2 shows the definition of the Davies-Bouldin index with four one-dimensional clusters.

### E. Training the classification object and testing

Based on the DBI values from the training section of the experiment, features are ranked and sorted in order of importance. The indexes of the most valuable 4000 features are saved, and these features used for training the LDA classification object. During testing, Gabor features from the new test epoch with the same indexes saved during training are used to test the classifier.

| | |
|---|---|
| $DBI = \frac{1}{4}\sum_{i=1}^{4} R_i$ | DB index, four classes |
| $R_i = max_{i \neq j}\left(R_{ij}\right)$ | Conservative approximation |
| $R_{ij} = \frac{S_i + S_j}{D_{ij}}$ | DB index, two classes |
| $D_{ij} = \sqrt{\left(c_i - c_j\right)^2}$ | Euclidean distance of centroids |
| $S_i = \frac{1}{N}\sum_{n=1}^{N}(x_n - c_i)$ | Standard deviation of class "i" |

***Equation 2.*** *Definition of the Davies-Bouldin index for 4 one-dimensional clusters. Most valuable features have the smallest DBI.*

## III. Results

### A. Classification Accuracy

The performance of the six participants are presented in Table 1. In average, all users got more than 16 out of 20 correct predictions (average performance 82.5).

### B. Time, Frequency, and Location of best features

From 6 users, and 4000 features/user, a total of 24K most valuable features were identified. The cumulative plots of the joint-time-frequency feature space, and the topographic map of the brain using these 24K features are presented in Fig. 2. Here, the top of the plot is the front of the head.

## IV. Discussion

As seen in Fig. 2, the key activity begins ~100ms post-onset. This corresponds with the time of Auditory cue recognition. The highest concentration is in the 70-128Hz range, which corresponds with Phonetic Linguistic activities.

The most important brain regions are the Prefrontal Cortex (related to trigger-driven executive control, waiting for time cue to begin task), Wernicke's area (related to Phonological code retrieval), and right IFG and Broca's area (related to Syllabification). The spatial, temporal, and spectral properties of the 24K most valuable features identified from 6 participants (Fig. 2), correspond to the automatic linguistic processing stages of word production prior to articulation, and are supported by a substantial body of evidence [2, 6, 10-13, 20, 21, 31-35].

The only conscious effort required from the user to operate this BCI is paying attention and waiting for the Auditory cue to covertly speak the desired command. All brain activities after auditory cue recognition (Phonetic Linguistic functions) are performed automatically by the brain and cannot be modified by the user. In fact, the 312ms trials end before conscious activity of any sort can begin in the brain.

The Linguistic BCI presented in this work is just as fast as a traditional P300 system. However, unlike the P300, this novel Linguistic BCI does not require constant gaze and attention to operate correctly.

***Table 1**: Classification Accuracy for all 6 users.*

| | Classification Accuracy |
|---|---|
| User 1 | 85% |
| User 2 | 80% |
| User 3 | 75% |
| User 4 | 90% |
| User 5 | 85% |
| User 6 | 80% |
| Mean | 82.5 ± 4.1% |

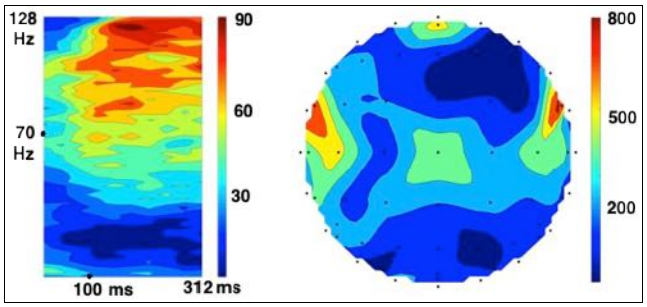



***Figure 2. (Left)** The cumulative colour-coded joint time-frequency representation of 24K best features from all users. The most important activity begins ~100ms post-onset corresponding to Auditory cue recognition and is concentrated in the 70-128Hz frequency range corresponding to Linguistic functions. **(Right)** The topographical plot of the brain using same features. The top of the plot is the front of the head. Most important regions: Prefrontal Cortex, left STG (Wernicke's area), right, and left IFG (Broca's area). These correspond with Linguistic Phonetic activity prior to articulation.*

In an earlier publication by these authors [8], a similar Linguistic BCI to this work, was compared with a Motor Imagery system, both using 1-second trials and the same experimental protocol and analysis pipeline. In an identical environment, the Linguistic BCI significantly outperformed the MI system. In addition, if trials are shortened from one second to 312ms, there is no significant loss of performance in the Linguistic BCI [9]. It is impossible to use MI for such a short trial duration.

For operating this novel BCI we used only those Linguistic functions that are completely automatic (Linguistic Phonetic functions) and temporally consistent. It is difficult to maintain perfect consistency for MI tasks, even during the same experiment. Considering these factors, Linguistic BCIs can potentially render MI systems obsolete.

In the present work, we only use a fraction of the time-frequency window containing Linguistic activity (up to 170Hz lasting over 2000ms) to successfully classify 4 covert speech tasks. By expanding the time and frequency of analysis, the number of word classes may be increased in future studies. Finally, the performance of the Linguistic BCI may improve by increasing the number of trials used for training [9] with the downside of increasing training duration and greater user fatigue.

## V. References

[1] E. Kraft, B. Gulyas, and E. Poppel, "Neural Correlates of Thinking," 1 ed.: Springer-Verlag Berlin Heidelberg, 2009, pp. 65-139.

[2] P. Indefrey and W. J. M. Levelt, "The spatial and temporal signatures of word production components," *Cognition,* vol. 92, pp. 101-144, 2004.

[3] E. C. Leuthardt *et al.*, "Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task," *Front Hum Neurosci,* vol. 6, p. 99, 2012.

[4] E. Kaan, "Event related potentials and language processing: a brief overview," *Language and linguistic compass,* vol. 1, no. 6, pp. 571-579, 2007.

[5] L. Y. Ganushchak, I. K. Christoffels, and N. O. Schiller, "The use of electroencephalography in language production research: a review," *Front Psychol,* vol. 2, p. 208, 2011.

[6] A. Llorens, A. Trebuchon, C. Liegeois-Chauvel, and F. X. Alario, "Intra-cranial recordings of brain activity during language production," *Front Psychol,* vol. 2, p. 375, 2011.

[7] A. Jahangiri and F. Sepulveda, "The contribution of different frequency bands in class separability of covert speech tasks for BCIs," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 2093-2096: IEEE.

[8] A. Jahangiri, J. M. Chau, D. R. Achanccaray, and F. Sepulveda, "Covert Speech vs. Motor Imagery: a comparative study of class separability in identical environments," in *Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE*, 2018: IEEE.

[9] A. Jahangiri and F. Sepulveda, "The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data," *Journal of Medical Systems,* vol. 43, no. 20, 2018.

[10] A. Flinker, E. F. Chang, H. E. Kirsch, N. M. Barbaro, N. E. Crone, and R. T. Knight, "Single-trial speech suppression of auditory cortex activity in humans," *J Neurosci,* vol. 30, no. 49, pp. 16643-50, Dec 8 2010.

[11] O. Creutzfeldt, G. Ojemann, and E. Lettich, "Neuronal activity in the human lateral temporal lobe," *Experimental BrainResearch,* vol. 77, pp. 451-475, 1989.

[12] M. Fukuda, R. Rothermel, C. Juhász, M. Nishida, S. Sood, and E. Asano, "Cortical gamma-oscillations modulated by listening and overt repetition of phonemes," *Neuroimage,* vol. 49, no. 3, pp. 2735–2745, 2010.

[13] V. L.Towle *et al.*, "ECoG gamma activity during a language task: differentiating expressive and receptive speech areas," *Brain,* vol. 131, pp. 2013-2027, 2008.

[14] M. A. Pitts, J. Padwal, D. Fennelly, A. Martínez, and S. A. Hillyard, "Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness," *Neuroimage,* vol. 101, pp. 337-350, 2014.

[15] F. Darvas, R. Scherer, J. G. Ojemann, R..P. Rao, K. J. Miller, and L. B. Sorensen, "High gamma mapping using EEG," *Neuroimage,* vol. 49, no. 1, pp. 930-938, 2010.

[16] S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations," *Frontiers in Human Neuroscience,* vol. 7, p. 138, 2013.

[17] J. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Frontiers in Human Neuroscience,* vol. 3, p. 61, 2009.

[18] I. BioSemi, "ActiveTwo-Multichannel, DC amplifier, 24-bit resolution, biopotential measurement system with active electrodes," 2001.

[19] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods,* vol. 134, no. 1, pp. 9-21, Mar 15 2004.

[20] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," *Front Neurosci,* vol. 5, p. 65, 2011.

[21] J. D. Greenlee *et al.*, "Human Auditory Cortical Activation during Self- Vocalization," *PLoS One,* vol. 6, no. 3, pp. 1-15, 2011.

[22] X. Chi, J. B. Hagedorn, D. Schoonover, and M. D'Zmura, "EEG-Based Discrimination of Imagined Speech Phonemes," *International Journal of Bioelectromagnetism,* vol. 13, no. 4, pp. 201-206, 2011.

[23] G. Gómez-Herrero, *Automatic artifact removal (AAR) toolbox v1. 3 for MATLAB.* 2007.

[24] G. Gomez-Herrero *et al.*, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," in *Signal Processing Symposium NORSIG*, 2006, pp. 130-133.

[25] C. Xun, H. Chen, and P. Hu, "Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis," *Journal of Applied Mathematics,* vol. 2014, pp. 1-10, 2014.

[26] Q. Shie and C. Dapang, "Optimal biorthogonal analysis window function for discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 42, no. 3, pp. 694-697, 1994.

[27] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Transactions on Signal Processing,* vol. 41, no. 7, pp. 2429-2438, 1993.

[28] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans Pattern Anal Mach Intell,* vol. 1, no. 2, pp. 224-7, Feb 1979.

[29] J. C. Rojas-Thomas, "New Version of Davies-Bouldin Index for Clustering Validation Based on Cylindrical Distance," in *V Chilean Workshop on Pattern Recognition*, 2013, pp. 81-86.

[30] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 12, pp. 1650-1654, 2002.

[31] Alycia Cummingsn, Amebu Seddoh, and Brianna Jallo, "Phonological code retrieval during picture naming: Influence of consonant class," *Brain Research,* vol. 1635, pp. 71-85, 2016.

[32] S. Chakrabarti, H. M. Sandberg, J. S. Brumberg, and D. J. Krusienski, "Progress in Speech Decoding from the Electrocorticogram," *Biomed Eng Lett,* vol. 5, pp. 10-21, 2015.

[33] P. Indefrey, "The spatial and temporal signatures of word production components: a critical update," *Frontiers in Psychology,* vol. 2, no. 255, pp. 255-271, 2011.

[34] N. S. Dhanjal, L. Handunnetthi, M. C. Patel, and R. J. S. Wise, "Perceptual Systems Controlling Speech Production," *The Journal of Neuroscience,* vol. 28, no. 40, pp. 9969-9975, 2008.

[35] J. Numminena and G. Curio, "Differential effects of overt, covert and replayed speech on vowel- evoked responses of the human auditory cortex," *Neuroscience Letters,* vol. 272, pp. 29-32, 1999.