Towards Phone Classification from Imagined Speech Using a Lightweight EEG Brain-Computer Interface

Jonathan Clayton

Masters' Dissertation
Speech and Language Processing
School of Informatics
University of Edinburgh

2019

Abstract

Along with my collaborator, Scott Wellington [1], we present the FEIS (Fourteen-channel EEG with Imagined Speech) dataset [2], a publicly-available dataset for use in investigating imagined speech recognition [1]. It consists of brainwave recordings from 22 participants collected using a lightweight, commercially-available 14-channel EEG device (the Emotiv Epoc+), in addition to the participants' audio recordings. The database is designed for developing EEG-based phone classification and regression models, and contains English- (21 recordings) and Chinese-language data (2 recordings). We benchmark our dataset against an existing dataset containing EEG recordings (KARA-ONE) [3], which, unlike ours, used a more standard 64-channel EEG cap.

Unlike [1], who focuses on regression models [2], this paper investigates phone classification. We compare three different data recording conditions: heard speech, imagined speech, and spoken (out loud) speech, as well as investigating different phonetic features (such as consonants/vowels and the voiced/voiceless distinction). We also compare the efficacy of using different machine learning models to classify the data (SVMs and CNNs), as well as different feature representations.

[1] The database is accessible at https://doi.org/10.5281/zenodo.3369179

[2] The author collaborated with Scott Wellington on data collection; we each worked separately on all other aspects of our respective investigations.

Acknowledgements

Thank you to my supervisors for their invaluable advice, to my collaborator, without whom I very literally could not have completed the project, and to everyone who participated in the data collection and endured hours of listening to monotonous recordings of their own voices - your help is very much appreciated!

Table of Contents

1. Introduction

Imagined speech recognition using brain signals is an area which has shown rapid advances in recent years, including notably [4] who used intracranial electrodes to synthesize speech. A key disadvantage of EEG compared with intracranial electrodes is the signal dampening and spatial smoothing caused by brainwaves being conducted through the skull [5]. However, a great advantage of EEGs compared to these methods is their non-invasive nature and ease of use [6]. Whereas most EEG devices designed for clinical applications use electrolyte adhesive gel to allow ideal scalp conductivity and accurate placement [7], other more lightweight, commercially available devices allow for being worn and removed easily.

In potential future applications, such as for use as a communication aid by people suffering from neurodegenerative disorders such as Motor Neurone Disease [8] [9], the characteristics such lightweight kinds of EEG are therefore obviously desirable. For this reason, we are interested in testing their feasability for these applications.

2. Background

This section gives background information about Electroencephalography (EEG) in general and Brain Computer Interfaces in section 2.1, a discussion of theoretical issues with machine learning from EEG signals in section 2.2, and a description of previous approaches to imagined speech recognition in section 2.3.

2.1 Electroencephalography (EEG)

Electroencephalography is a brain-imaging technique used for measuring electrical potentials generated by brain structures. It is a non-invasive technique; potentials are measured by electrodes placed on the scalp [10].

2.1.1 Comparison to other methods of brain imaging

Compared to techniques such as fMRI (functional Magnetic Resonance Imagery), which detects differences in levels of blood flow within the brain, EEG has a poor spatial resolution. However, the temporal resolution of EEGs is far superior, since the temporal resolution of MRI is intrinsically limited by the speed of the hemodynamic response (or blood flow in the brain) [11].

EEG, relying on electrical potentials, has a higher temporal resolution. Many EEGs have a sampling frequency of around 1000 Hz (e.g. [3][12][13]). However, the periodic signals used for classification of motor imagery (imagined movements) using EEG tend to be of still lower frequency - other researchers use devices with a sampling frequency on the order of 256Hz. Owing to the success of these approaches, it seems that imaging only brainwaves with a frequency of below the Nyquist frequency of 128 Hz is sufficient for recognition (e.g. [14] [15]).

Another brain imaging technique which could potentially be used for speech recognition from brainwaves is MEG (Magnetoencephalography), which has a similar temporal resolution to EEG [6]. However, since MEG devices tend to be much larger than EEGs, their use is not currently feasible outside of a laboratory setting, and they are hence less suitable as Brain-Computer Interface (BCI) devices [16].

2.1.2 Brain-Computer Interfaces (BCIs)

One device, currently in use, that uses EEG signals for a brain-computer interface (BCI) is the P300 speller[17] [18], relying on the P300 "oddball" response, a spike in brainwaves generated by surprise or novelty . P300 spelling systems work by presenting users with a screen showing a 2D matrix of letters, from which a single letter must be selected. While the user concentrates on a letter, different rows and then columns are highlighted randomly as a bar repeatedly flashes on the screen. The P300 response is generated when the column or row containing the letter that the user is focusing on is flashed. While these devices are judged by some researchers as being the most effective BCIs currently on the market, a disadvantage is the slow typing speed possible with the devices, which is in the order of 5-10 characters per minute [19]. Even though spelling can somewhat be sped up using language modelling, typing speed is still far from the average 120 *words* per minute produced during fluent speech[20]. Hence, there is a strong motivation to develop alternative BCI devices which may improve the speed of communication.

2.1.3 EEGs and electrical potentials

EEG devices are thought to record two types of electrical potentials; action potentials and post-synaptic potentials. Action potentials are the primary way which neurons in the brain communicate. They are initiated when positively charged sodium (Na+) ions build up outside the soma (cell body) of a neuron. When the Na+ ions reach a certain threshold, ion channels are opened in the cell, causing a feedback loop in which a rapid influx of Na+ ions is triggered. This sudden positive charge is then transmitted down the axon (tail) of the neuron. Hence, there is a sudden "spike" of electrical potential inside the cell [21].

Although these potentials may be most relevant for understanding brain dynamics, they are hard to measure using EEGs since action potentials often cancel each other out. Hence, they can only be detected when there are many neurons "firing" at once. What the EEG mostly measures are the second type of potential post-synaptic potentials, which correspond to the build-up of Na+ ions described in the previous paragraph. These are easier to detect than action potentials due to their longer duration [22].

2.1.4 EEG devices and the Emotiv Epoc+

Standard EEG devices often use 64 electrodes, mounted on the skull using electrolyte gel. They are standardly positioned according to the international 10-20 system [23], shown in figure 2.1.

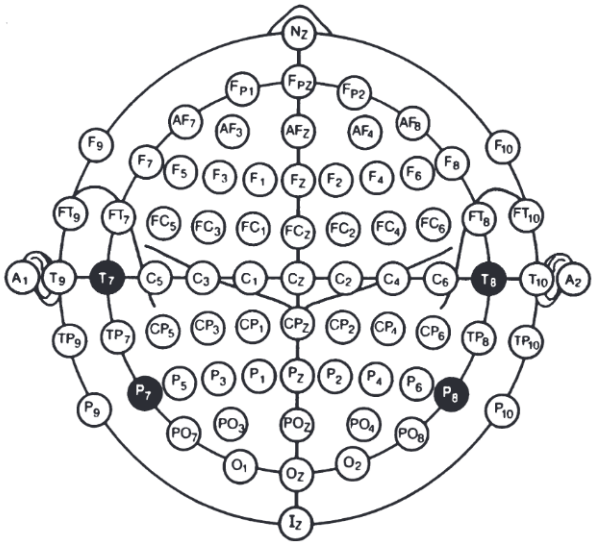


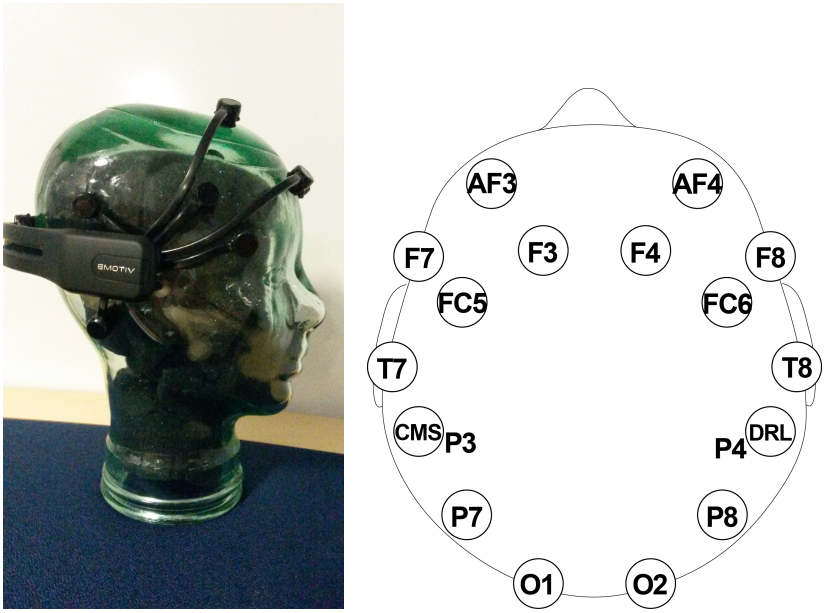
Figure 2.1: Positioning of electrodes according to the international 10-20 system (from [23]). This is a top view of the head, the nose is at the front.

The Emotiv Epoc+, used in this experiment, uses 14 copper electrodes with felt pads which are held in place with a plastic mounting [24]. This makes maintaining the same relative positioning of electrodes easy (within subjects), making it ideal as a device to be placed and removed easily during regular wear. However, assuming the device will be worn by people with different head shapes, this characteristic also makes it impossible to maintain, between subjects, a consistent positioning on the skull. This is something that may make subject-independent recognition using the device harder, since spatial information is taken into account by all the models we are testing here.

The Emotiv Epoc+ has a maximum sampling frequency of 256Hz. It uses 14 electrodes, whose average positioning on the human skull (according to the international 10-20 system) is at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 [23]. Two reference (ground) electrodes are used to help filter out electromagnetic interference from outside the body (e.g. from power lines). These are positioned at P3 and P4 [24].

2.2 Issues with machine learning from EEG data

EEG signals pose a challenge for machine learning for at least four reasons; firstly, the data is high-dimensional; secondly, the data contain a large number of artifacts; and thirdly, the sources of informative data tend to be highly mixed. Fourthly, there is a large degree of variation between subjects that machine learning models need to account for.



(a) (b)

Figure 2.2: The Emotiv Epoc + device, and a map of electrode positions on the scalp according to the international 10-20 system (adapted from [24])

2.2.1 Dimensionality of EEG signals

The high-dimensionality of the data stems from the fact that we are dealing with a three-dimensional surface (though treated as 2 or 1-dimensional by the models used here) that changes across time. All the models we are using take into account temporal information, which is fundamental for brain-based machine learning tasks. There is evidence that EEG signals are hierarchical in the temporal domain (e.g. [25], [26]).

2.2.2 EEG artifacts

Artifacts in EEG come from multiple sources. Firstly, there are artifacts external to the body such as electomagnetic interference, which are ameliorated by the ground (reference) electrodes used by the Emotiv device [24]. Second, and more significantly, there are signals caused by muscle movements (electromyography or EMG), eye movements (electrooculography or EOG) and heartbeats (electocardiography or ECG). A commonly-used technique for attempting to eliminate these artifacts is ICA (Independent Component Analysis) [27]. We attempted an implementation of ICA (described in section 4.2.2), but found that it worsened baseline performance, so the models we chose instead used unprocessed data.

2.2.3 Mixing of sources of electrical potentials

A third problem is the mixing of different components. Matter in the brain, skull and scalp, through which EEG signals are conducted before reaching electrodes, works to spatially smooth the signal - hence, signals from multiple sources in the brain are mixed when they arrive at electrodes. Therefore, learning from EEG signals is somewhat similar to the "cocktail party problem"[28] in which we attempt to isolate the voices of different individuals in a party using a set of recordings from multiple microphones; in the brain, neuron ensembles are analogous to different speakers, and the electrodes are analogous to microphones. Our models must attempt to separate this information in order to learn well.

This is another problem which ICA could potentially help to ameliorate. CNN models may also learn an "unmixing matrix" similar to the one calculated in ICA [29].

2.2.4 Inter-subject variation

Much variation seems to exist between EEG signals for different subjects. At least one source of this is the differing shapes of people's skulls and brains [30]; regardless of whether or not similar brain activity is taking place, being filtered through different individuals' cerebral cortices and skulls will likely alter patterns of electrical activity in ways that are hard to predict. Patterns of brain activity between individuals are unique enough that subject identification using EEG signals has been proposed [31].

These factors make subject-dependent classification a much easier task than subject-independent classification. Subject-independent classification is of course desirable since it would make setting up BCIs for new individuals a significantly less time-consuming task in the future. However, subject dependent classification is likely to result in significantly higher accuracies.

2.3 Previous approaches to imagined speech recognition

In this section, we discuss an already-existing dataset containing EEG recordings, the KARA-ONE dataset (section 2.3.1), followed by a short survey of machine learning models used by the KARA researchers and others for imagined speech recognition (section 2.3.2).

2.3.1 The KARA-ONE dataset

The KARA-ONE dataset is an open source dataset of multi-modal speech recordings (audio, EEG recordings, and facial recordings using Kinect software) [3]. The KARA-ONE data was recorded using an EEG that was less lightweight (and had higher spatial

resolution) than the one used in this dissertation - the 64-channel Neuroscan Quick Cap. Unlike the Emotiv Epoc+, this device requires the use of electrolyte gel. However, it is still fairly lightweight and quick to put on when compared to EEGs primarily designed for clinical applications, given that all the electrodes are pre-positioned inside a fabric cap, instead of requiring individual application to the skull by a researcher. The KARA researchers used two bipolar ocular channels (4 electrodes in total) above and below the left eye, and to the lateral side of each eye, to help remove blinking and eye movement artifacts.

The full KARA-ONE data collection procedure is summarized in (cite). The KARA-ONE dataset contains 7 phonetic/syllabic prompts (/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) as well as 4 whole word prompts, which we did not use in this study.

The basic structure of a single data collection "Epoch" (corresponding to a single sound or prompt) in their experimental procedure is was as follows:

1. A 5 second "clearing" phase - a blank screen, during which participants attempt to clear their mind.
2. A prompt is diplayed on a computer screen (the name of the sound) and simultaneously play a recording of the sound.
3. A 2 second "prepare articulators" phase, in which participants move their articulators into position, as if they were about to produce the sound in the prompt.
4. A 5 second "imagined speech" phase, in which participants imagine articulating the sound
5. A "speaking" phase, in which participants actually say the sound out loud.

The general recording structure of the EEG recordings in the KARA-ONE dataset is a useful one, since it allows for the testing of multiple experimental conditions ("thinking","hearing" and "speaking"). For our experiments, we maintain the same general structure with some minor modifications (described in section 4.1.5).

2.3.2 Previous approaches to speech recognition from brainwave data

A wide variety of approaches to speech recognition using brain data have been tried before, on various devices and using various machine learning models. On just the KARA dataset, researchers have used SVMs, deep belief networks, and hybrid CNN/LSTMs.

Zhao and Rudzicz use SVMs as their baseline model. SVMs may be a good choice of classifier for EEGs since they can work with an arbitrarily high number of features. Zhao and Rudzicz compute $19 \times 19 \times 62 = 22382$ features for every prompt in their dataset. As a simple dimensionality reduction technique, they calculate the correlation of each of these features with the phonetic feature being tested, and keep only the n most correlated features for use in classification.

Many approaches also use neural models. For example Zhao and Rudzicz also use a Deep Belief Network (DBN). In [32] Saha, Fels and Abdul-Mageed make improvements over the SVM in Zhao and Rudzicz's paper using a novel cross-correlation-based dimensionality-reduction technique, and a hybrid model which combines a CNN, an LSTM and a denoising auto-encoder.

One specific class of methods which seem promising for use with EEG data are convolutional neural networks (CNNs) [33]. One key benefit of CNNs could be in dimensionality reduction since EEG data are very high dimensional. CNNs are also designed to work with data containing hierarchical features, which EEG data may contain in the time dimension [34]. Some brain oscillations appear to be hierarchical (e.g. [25], [26]); this results from the phase of lower-frequency oscillators modulating higher-frequency oscillators. These hierarchical oscillations could be relevant to language processing [35].

2.3.3 Feature selection from EEG datasets

As described in the previous section, dimensionality reduction for SVMs is posisble using a simple correlation-based technique, which we replicate in this paper. Zhao and Rudzicz demonstrate that the number of features is crucial for optimizing models for classification accuracy [3].

The issue of how to input data into neural models is also likely to be critical. When using CNNs (Convolutional Neural Networks), at least two different feature representations are possible. One method, which is at least intuitively appealing, is to try to preserve the both the spatial relationships between electrodes as well as temporal information in the input. Bashivan et. al. do this by producing EEG "images"; they first project the three-dimensional locations of electrodes on the skull into a two-dimensional plane (using an Azimuthal Equidistant Projection - the map projection used in the logo of the United Nations). Interpolation is used to produce a 32×32 matrix [36].

The disadvantage of this technique is that it is extremely high-dimensional. To reduce the dimensionality somewhat, Bashivan et. al. window their data and use a Fast Fourier Transform to split it into frequency bands (Theta: 4-8Hz, Alpha: 8-12Hz and Beta: 12-30Hz). The shape of the final data, therefore, is 4D: n time windows $\times$ n. channels $\times$ image width $\times$ image height. This results in a representation analogous to an RGB video file format.

Another, simpler method allows raw data to be used in the input representation; this is a 2D matrix with a size of n. channels $\times$ n. timesteps. The disadvantage here is the loss of spatial information, however, an advantage is the fact that spectral information can be learnt implicitly from the data, rather than being hard-coded as in the case of a "video-style" input [29].

It is an open question as to which representation of the data is better for obtaining optimal classification accuracies; Schirrmeister et al. [29] argue that a 2D representation of spatial information in the input is unnecessary due to the spatially-smoothed

nature of EEG inputs. Conversely, Bashivan et al. [36] state that a potential advantage of the video technique is that input from different devices can be converted to the same invariant video-style representation (by the projection and interpolation method described above).

Another novel alternative feature representation uses cross-covariance matrices as input. The utitily of these is as a dimensionality-reduction technique. This reportedly achieved high accuracies on the KARA-ONE data [32].

3. Overall Hypothesis and Research Questions

Our overall hypothesis is that a more lightweight device (the 14-channel Emotiv Epoc+) can provide comparable results on sound categorization tasks to a more heavy-duty model (the 64 channel neuroscan Quick Cap).

In addition to this, we investigate five research questions:

- Do CNNs perform better than SVMs for sound categorization?
- Which features (for SVMs) / feature representations (for CNNs) are best for sound categorization with EEGs?
- Which phonetic features are easiest / hardest to detect using an EEG?
- Which experimental phase (hearing, imagined speech, or actual speech) is best for sound categorization, and what differences are there between them?

4. Methods

4.1 Data Collection

We collected data from 22 participants. From these participants, we collected 21 recordings of English language data, and 2 recordings of Chinese language data (Participant 15 in the English recordings is Participant 01 in the Chinese recordings).

4.1.1 Accessing brainwave data from the EEG device

We used the open-source software package OpenVibe [37] to interface with the device. The software contains an acquisition server, which communicated with the wireless USB dongle which comes with the Emotiv Epoc+. We also used OpenVibe Designer to create experimental scenarios. This allowed us to control the flow of the experiment by showing prompts at consistent time intervals, as well as giving us data that was "Epoched" into sections corresponding to the lengths of time that prompts were displayed onscreen.

4.1.2 English language data

We collected data from 21 participants, all native speakers of English or fluent non-native speakers, with no known neurological disorders. Three of the participants (participants 03, 09 and 19) are left-handed and one (participant 08) is ambidextrous. All of the other participants are right-handed. We obtained ethical approval for the experiment, and all participants gave their consent to participate. Each recording session took approximately 90 minutes, including making sound recordings, setting up the device, and making EEG recordings.

4.1.3 Chinese language data

We collected data from 2 participants, both native speakers of Chinese, with no known neurological disorders. Ethical procedure was identical to that for the English recordings. We chose to make some recordings in Chinese, a tonal language, as well as

English, since this allows us to investigate how well phonetic tone can be detected by EEG recordings (albeit only in a tentative way, due to the low number of participants).

4.1.4 Voice recordings

Participants were played recordings of their own voices during the task. Prior to collecting brainwave data, we made recordings of each participant's phonemes, using a DPA 4088 cardioid microphone in a hemi-anechoic chamber. Phonemes were produced in isolation, with the exception of plosives, for which participants were instructed to produce a schwa after each phone, e.g. /kə/ /pə/ /tə/, to aid participants with distinguishing the phonemes.

4.1.5 Outline of experimental procedure

Participants were presented with a total of 16 categories $\times$ 10 repetitions = 160 prompts. These were presented to participants in a random order. As in KARA-ONE, there were five phases of the experiment which were repeated for each prompt. However, unlike KARA, we did not instruct participants to prepare their articulators in the pause between the "stimuli" and "thinking" stages. This was to avoid potentially introducing muscle movement artifacts which could be confounded with the imagined speech data.

Participants were instructed to keep as still as possible and to limit eye movement and blinking during all phases of the experiment with exception of the "clearing" phase. The phases of the experiment were as follows:

In the first phase, participants were instructed to clear their mind for five seconds; next, they listened to five repetitions of the recorded phoneme while the IPA symbol for that phoneme was displayed on the screen; next, there was a one-second pause; next, participants imagined saying the phoneme five times; finally, participants said the phoneme five times aloud.

The duration of the EEG recording parts of the experiment was 60 minutes for each participant. Breaks were provided to allow participants to rest and not lose concentration on the task; recording was divided into three sections lasting 23.5, 19.75 and 19.75 minutes respectively.

4.1.6 Fitting the EEG device

Prior to fitting the EEG device, we first removed the felt pads covering the electrodes and saturated them with saline solution, before replacing them in the device. This measure improves conductivity from the electrodes to the scalp. We then placed the device on the head. To attempt to maintain a consistent positioning of the device between subjects, we placed the two unused rubber-tipped reference electrodes at the bottom of the device on the mastoid process (a bony protrusion of the skull behind

	bilabial/labiodental	alveolar	postalveolar/palatal
voiceless plosive	p	t	k
voiceless fricative	f	s	sh
voiced fricative	v	z	zh
voiced nasal	m	n	ng

Table 4.1: Consonants

	Front	Back
High	i	u
Low	æ	ɔ

Table 4.2: Vowels

the ear), and then rotated the device until electrodes AF3 and AF4 were at a distance of three of the researcher's fingers from the participants' eyebrow ridge. Due to the fixed shape of the device, consistent positioning of electrodes between subjects with different head shapes was not always possible. We used the Emotiv BCI to check conductivity of the device with the skull. This application contains a tool to check conductivity, which we ensured was always at 100% for each participant prior to the recordings. Where required, we worked electrodes between participants' hair in order to obtain a better connection.

4.1.7 Prompts used

For the English recordings, we used 16 English phonemes. We chose this set so that there was a reasonable variety of phonemes (and therefore also quantity of data in each category) when we split the dataset between various phonetic categories (e.g. vowel/consonant, voiced/voiceless, plosive/fricative/nasal, labial/alveolar/palatal, etc.).

For the Chinese recordings, we used 16 syllabic prompts, all of which corresponded to several real Chinese characters (either words or syllables).

Their pronunciations in Pinyin (the standard Chinese phonetic script) are shown in the table below:

First tone (high)	mā	mēng	duō	tuī
Second tone (rising)	má	méng	duó	tuí
Third tone (falling/rising)	mǎ	měng	duǒ	tuǐ
Fourth tone (falling)	mà	mèng	duò	tuì

Table 4.3: Chinese

These prompts allowed us to do both phoneme-based and tone-based categorization using Chinese data, with four equally-sized categories, and including a large degree of phonemic variation.

4.2 Pre-processing the data

Pre-processing of the data took place in various stages of our model-building pipeline. The pre-processing performed by the device itself is described in section 4.2.1. Our tests with ICA artifact removal (which we did not eventually apply to the data) are described in section 4.2.2. The pre-processing steps applied for each model individually are described in sections 4.2.3-4.2.6.

4.2.1 Pre-processing by the Emotiv Epoc+

The Emotiv Epoc+ uses 50 and 60 Hz digital notch filters to remove electromagnetic interference from power cables [24].

4.2.2 Artifact Removal and ICA

Prior to building the models, we tested the possibility of using ICA (independent components analysis) for artifact removal and to separate different sources of electrical potentials within our brain data (Blind Source Separation). The KARA-ONE data had two bipolar EOG (electroculography) channels to aid with removal of artifacts - HEO (Horizontal) and VEO (vertical) channels.

For the KARA data, our artifact removal/ ICA process had four stages, which we repeated for each speaker:

1. Perform ICA (using the FastICA algorithm) on the raw data, obtaining an unmixing matrix.
2. Calculate correlations of ICA components with the HEO/VEO channels.
3. Eliminate highly correlated components of the ICA, and use the unmixing matrix to calculate the components to be retained.
4. Perform ICA again on the cleaned data, this time without eliminating components, to obtain (hopefully) meaningful components corresponding to sources of potentials in the brain.

We carried out this process using the implementation of the FastICA algorithm available in the MNE python package for EEG data processing [38], following the tutorial on the MNE website [39].

The Emotiv Epoc+ lacks EOG channels for removing ocular artifacts. However, it may be possible to use an alternative technique to carry out ICA; one could take advantage of existing ICA solutions such as the ones calculated on the KARA-ONE data and carried out then perform the same procedure described above. In this case, instead of calculating correlations of each component with the HEO/VEO channels in step 2, we would calculate correlations of each Emotiv ICA component with the “bad” components of the KARA ICA solutions [39].

Feature	Without ICA	With ICA
C/V (thinking)	71.6	69.6
C/V (hearing)	65.2	62.9
Voiced/voiceless	69.2	71.0
Average	68.5	67.8

Table 4.4: Average % classification accuracy on the 5-fold validation sets, using an SVM, using a speaker-independent model.

We checked whether our ICA techniques were effective by running our baseline model (an SVM) on data which had not been preprocessed with ICA, and compared mean cross-validation scores on the pre-processed data vs the un-preprocessed data. The results are shown in Table 4.4.

These scores indicate that performing ICA did not result in a significant improvement in classification accuracy - if anything, it seems to slightly worsen classification. For this reason, we did not use ICA on either the KARA or FEIS datasets.

In order to investigate why ICA did not appear to be effective, we carried out a visual inspection of the plots of ICA components generated by the MNE package. None of the components plotted, including the ones that had been found to be correlated with EOG channels, had the characteristic dark patches near the eyes associated in these plots with EOG artifacts [39]. This would seem to imply that in this case the ICA algorithm has failed to identify these components.

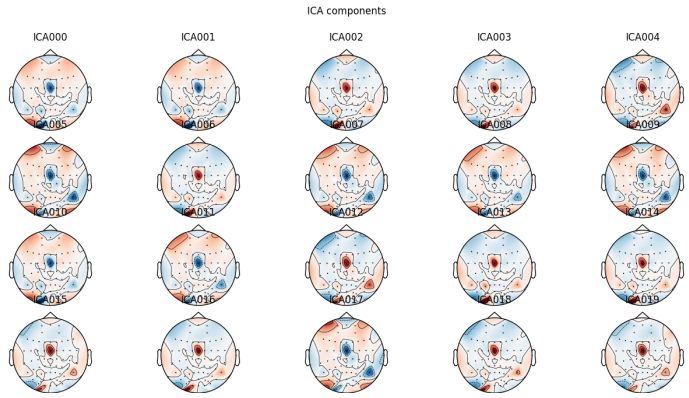


Figure 4.1: Plots of estimated sources of ICA components generated by the MNE Python package. [38]

Investigating alternative implementations of ICA, or artifact removal algorithms, could be a fruitful avenue of future research, but goes beyond the scope of this investigation.

4.2.3 Pre-processing for the SVM

Our implementation of the SVM was based on Zhao and Rudzicz. Zhao and Rudzicz first compute a large series of statistical features for every Epoch of brain data, and then

use a correlation-based dimensionality reduction technique to select the most relevant features for the phonetic contrast in question [3].

Features are calculated independently for each electrode. We first window each Epoch into 19 square windows, with 50% overlap between windows (three windows of previous context are also needed for the calculation of the delta, double-delta features and the Enhancement Factor) (for Enhancement Factor see [40]).

We calculated 28 mathematical features for each window, 19 of which were taken from Zhao and Rudzicz. Additionally, we use implementations of entropy, fractal dimensions, and other time series features from the python package EntroPy [41]. The full list is shown in table 7.1 in the appendices. For each of these features, we also calculated Delta and Delta-Delta features, so for every Epoch (prompt) we calculated 3×28 features $\times$ 19 windows $\times$ 14 electrodes = 22344 features for our data, or, for the KARA data, $\times$ 62 electrodes = 98952 features.

4.2.4 Pre-processing for the video-recognition style CNN

We used Bashivan et al's Lasagne implementation of a video-recognition style CNN [36].

As for the SVMs, we used 19 windows. In this case, we used Hanning windows instead of square windows, as we subsequently applied a Fast Fourier Transform to obtain the power spectrum, which we binned into Theta (4-8Hz), Alpha (8-12Hz), and Beta (12-30Hz) bands.

We used interpolation (using Bashivan et. al's code) to obtain 32×32 matrices (or "images") for each window, making the overall shape of the data $19 \times 3 \times 32 \times 32$. This shape was consistent regardless of the dataset used (FEIS or KARA).

4.2.5 Pre-processing for the 2D-input CNN

Since the 2D-input CNN takes raw input, we did not perform any pre-processing on the data prior to feeding it into the model, with the exception of the pre-processing carried out internally in the Emotiv device, as described above in section 4.2.1.

4.2.6 Pre-processing for between-datasets testing

When we attempted training and testing on the above models using the SVM or the 2DCNN, we downsampled the KARA data from 1000Hz to 250Hz (to be closer to the 256Hz sampling frequency of the FEIS data) and normalized the FEIS data to have a mean of 0 (as the KARA researchers did on their data).

4.3 Sound Categorization

We compared three statistical models for sound categorization: an SVM, and two types of CNNs. For the SVM, we use 5-fold cross-validation, while for the CNNs we use 16% of the overall data as the validation set. We train and test independently on data from the three test conditions (speaking, hearing and thinking). We use both subject dependent and independent models. For the subject dependent models, we attempt training on just our data, on our data and the data from KARA-ONE combined, and just on the KARA-ONE data.

4.3.1 Test and training sets

In the data we have collected, the frequency of sounds in some phonetic classes (such as consonants) was higher than that of others (such as vowels). In order to avoid misleadingly high classification accuracies, we ensure that there are an equal number of training and test examples from each class. In addition, as far as possible, we attempt to include an equal number of sounds from each sub-class within the classes (for example, an equal number of every vowel type we recorded within the vowels class). This is to ensure that we are really classifying based the phonetic contrast under investigation; for example, in the case of consonants vs. vowels, that we are really distinguishing the two categories instead of just /t/ vs. /u/ for example.

When testing speaker-dependent models, our test data is 20% of the relevant tokens. When testing speaker-independent models, we tested our models on a single speaker at a time and trained on all other speakers.

4.3.2 Summary of models used

Abbreviation	Model Type	Implementation
SVM	Support Vector Machine	Scikit-Learn, based on description in Zhao and Rudzicz
VCCNN	Video-Classification style CNN	Bashivan et al
2DCNN	2-dimensional input CNN	Deep4Net, Braindecode

Table 4.5: Abbreviations for the classification models used in the report

4.3.3 SVM

We followed Zhao and Rudzicz in taking all of these features as a single vector, and then, for each classification task, calculating the Pearson correlation coefficients of each of these features with the linguistic feature of interest (for multi-class tasks, we

used point biserial correlation). As part of the parameter sweep for optimizing the SVMs, we chose the optimal N of most highly correlated features, in N $\in$ [5:100].

We used 5-fold leave-one-out cross-validation to tune the model, using an implementation in the Scikit-Learn Python package [42]. We followed Zhao and Rudzicz in using a Radial Basis Function kernel, and tuned for optimal values of C in the range [1:1000] and gamma in the range [0.001:0.000001], using $\log_{10}$ steps. We tuned for optimal F-score.

In addition to calculating classification accuracies for the SVM model, we also use the correlations we have calculated to try and answer the research question of which features, and electrodes of the device, are most correlated with the linguistic features we are testing (see sections 5.5-5.6).

4.3.4 CNNs with two-dimensional input

For the 2DCNNs, we did not do any feature engineering on the raw input before passing it into the model; hence the input had a 2D shape with time on one axis and electrodes (unordered) on the other.

As a representative of this class of model, we used the implementation of Deep4Net [29] in the Braindecode Python package [43]. This model was found to be effective across a range of classification tasks in [34].

To train this model, we used the Adam optimizer and a learning rate of 0.0625. For every task, we also performed a grid search to find optimal hyperparameter values for the number of filters in each layer n $\in$ [25, 50, 100, 200], pool/filter lengths n $\in$ [5, 10, 20, 40], and stride lengths n $\in$ [3, 6, 9, 12]. We used a dropout probability of 0.5 and Elu nonlinearities. We trained each model for 30 Epochs.

4.3.5 Video classification-style CNN

For the VCCNN, we used Bashivan et al's Lasagne-based implementation (EEGLearn) [36].

This model consisted of 19 parallel CNN ensembles (each containing a 4-layer, 2-layer and 1-layer model), followed by a max-pooling layer, a fully-connected layer and a softmax layer. Following Bashivan et. al, we trained each model for 6 Epochs [36].

4.3.6 Sanity Checking

Prior to collecting subject-independent results, we carried out a sanity check on our data. We did this by using our baseline model (the SVM) to carry out subject-dependent categorization. We took the average of three binary tasks across three modalities (thinking, speaking and hearing), obtaining nine results in all. We trained

on 80% of the data and calculated average validation scores for each task. The results are shown in Table 4.6.

Speaker ID	Average Validation Accuracy
19	84.53%
09	84.45%
11	83.48%
12	83.48%
20	83.40%
17	82.84%
05	82.29%
06	82.07%
15	81.46%
07	81.40%
16	81.14%
13	80.64%
18	80.54%
03	80.50%
08	78.48%
14	77.77%
21	76.97%
04	75.71%
10	74.26%
01	74.26%
02	72.97%

Table 4.6: Percentage validation accuracies for each participant across 3 speaker-dependent tasks.

As the table shows, the average validation accuracy for all speakers was between 72 and 85 %.

The low accuracy for participant 02 could be explained by the fact that the electrodes were moved by a substantial amount during recording. However, since the accuracy is not much lower than that for the other participants, we decide to retain participant 02's data.

5. Results and Discussion

In this results section, we first present subject dependent results for classification accuracy on the two datasets (sections 5.1-5.3) followed by subject independent results (section 5.4). We also present averaged correlations of the linguistic features we tested with EEG data features calculated from different electrodes on our device (section 5.5) as well as a list of averaged correlations of each EEG data feature with the linguistic features (section 5.6).

5.1 Comparison of Datasets (FEIS and KARA)

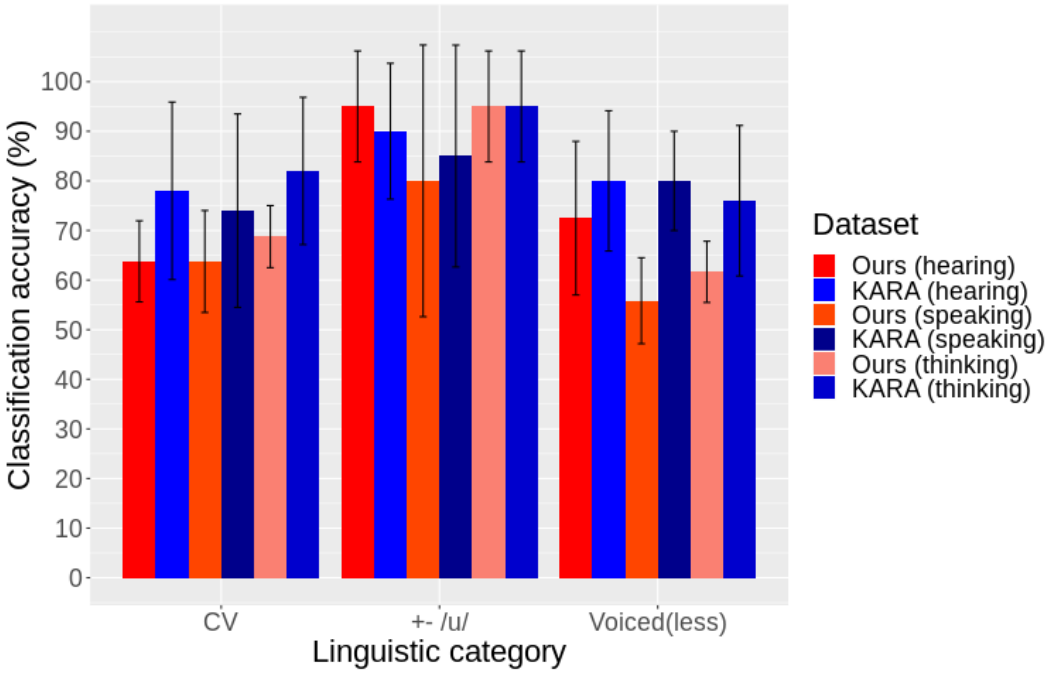


Figure 5.1: Mean subject dependent classification accuracy scores for three tasks across five randomly-selected speakers from the two datasets. Error bars are standard deviations.

The results in figure 5.1 above indicate that for most tasks, the data from the KARA dataset, collected using a 64-channel EEG, only resulted in a slightly better classifica-

tion accuracy than our data collected with a 14-channel headset.

In the C/V and voice tasks, in fact, these differences are less meaningful, since the C/V and voiced/voiceless may have be somewhat intrisically harder for our dataset. We recorded a greater variety of consonants/vowels and voiced/voiceless sounds, and there was therefore more internal variation within our dataset. For example, we recorded 6 voiced and 6 voiceless consonants, compared to KARA's 2 and 3. In the $\pm/u/$ task, where the variation was comparable, there appear to be no significant differences between classification on our dataset and on KARA.

5.2 Subject-dependent results (English data)

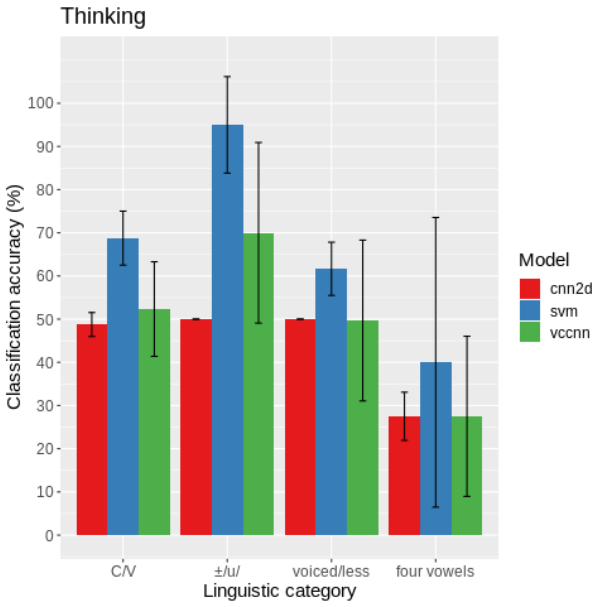
The graphs in figure 5.2 show subject dependent results, in which the model was trained and then tested on the same subject. It is clear from looking at the graph that the SVM was the best-performing model, as it had a higher average score for every linguistic category/ test condition combination. The other models do not seem to perform substantially better than chance, although the video-recognition style model appeared to perform somewhat better than the 2DCNN model in some cases, such as for categorizing $\pm/u/$ in the thinking condition, for which the average accuracy was 70%.

Out of the four phonetic contrasts, unsurprisingly $\pm/u/$was the highest-scoring task across the four, since only one sound had to be distinguished from the others, resulting in less inter-class variation in one of the two binary categories. Out of the two tasks which had more inter-class variation, the C/V task and the voiced/voiceless task, slightly better results were achieved for the C/V task in two of the three conditions. Interestingly, SVM results for voiced/voiceless were better (at 71.5% vs 61% / 55.5%) in the hearing task than in either the thinking or speaking tasks. We can speculate that in this case, the SVM took advantage of activations in Broca's area detected by the device, which may have been particularly strong in the hearing condition (see the averaged feature correlations with electrodes discussed in the section below).

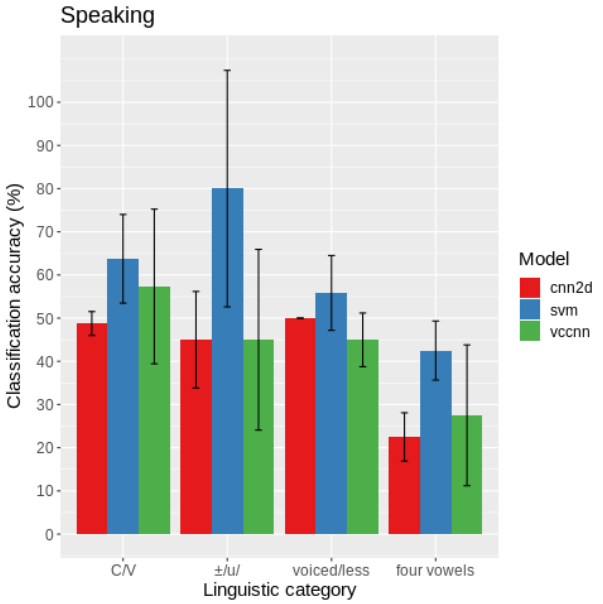
Comparing the three different test conditions, the best accuracies were achieved on the hearing data, with an average of 70.8 % across the four tasks with the SVM. The next most accurate was the thinking data with 66.4%, then the speaking data with 60.3%.

5.3 Subject-dependent results (Chinese data)

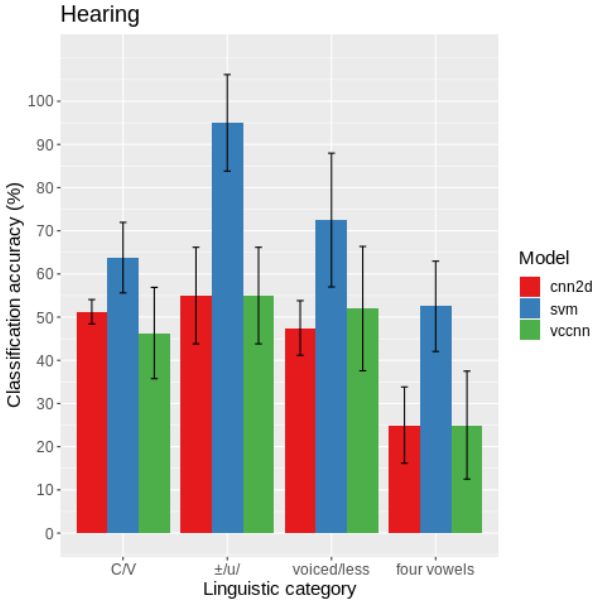
In order to test the ability of the Emotiv device to detect information that is useful for distinguishing phonetic tone, we ran our speaker-dependent SVM model on the data from our two Chinese-speaking participants. We did four-class classfication on categories of sounds with the same segmental phones (but varying tones) and categories with the same tones (but the varying phones). Our results are shown in table 5.1.



(a)

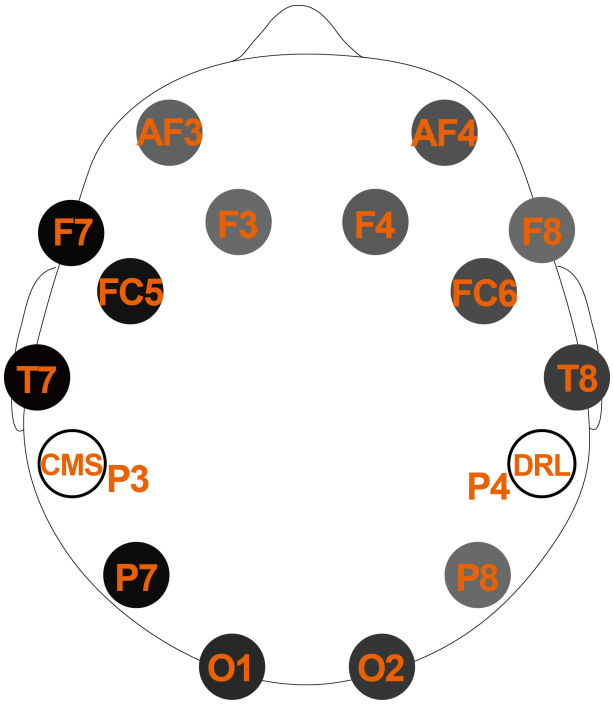


(b)



(c)

Figure 5.2: Subject-dependent results for classification accuracy on the FEIS data

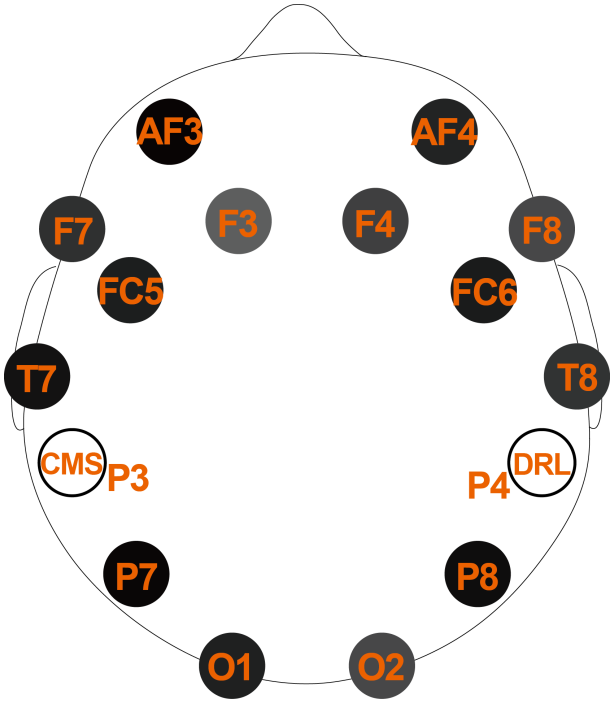


(a) Caption of subfigure 1

Electrode reference	Average Correlation
T7	2.81e-3
F7	2.77e-3
P7	2.67e-3
FC5	2.58e-3
O1	2.35e-3
O2	2.22e-3
T8	2.15e-3
FC6	2.01e-3
AF4	1.90e-3
F4	1.82e-3
AF3	1.74e-3
F8	1.67e-3
F3	1.67e-3
P8	1.66e-3

(b) Caption of subfigure 2

Figure 5.3: Averaged correlations with linguistic features for the thinking condition

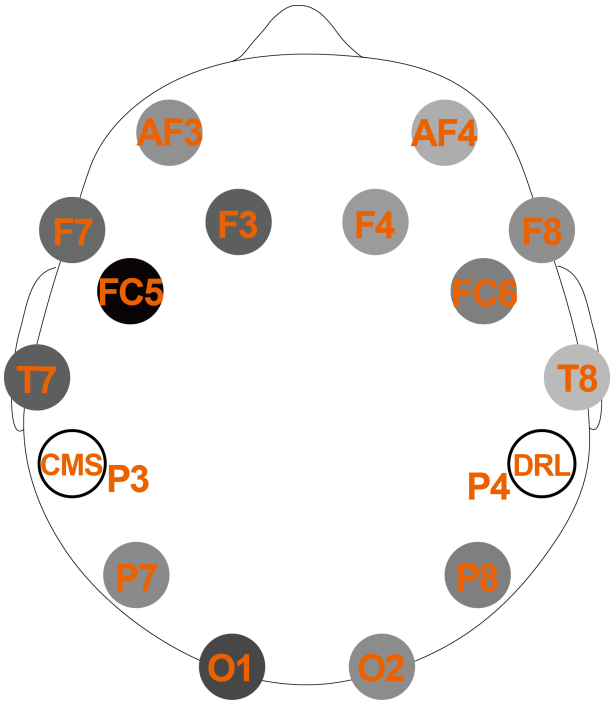


(a)

Electrode reference	Average Correlation
AF3	2.37e-3
P7	2.31e-3
P8	2.24e-3
T7	2.18e-3
FC6	2.11e-3
FC5	2.09e-3
O1	2.06e-3
AF4	2.03e-3
F7	1.91e-3
T8	1.90e-3
F4	1.77e-3
F8	1.75e-3
O2	1.70e-3
F3	1.48e-3

(b)

Figure 5.4: Averaged correlations with linguistic features for the speaking condition



(a) Caption of subfigure 1

Electrode reference	Average Correlation
FC5	2.01e-3
O1	1.45e-3
F3	1.27e-3
T7	1.26e-3
F7	1.19e-3
FC6	1.01e-3
P8	9.96e-4
P7	9.20e-4
O2	9.09e-4
F8	8.77e-4
AF3	8.53e-4
F4	7.91e-4 a
AF4	6.38e-4
T8	5.39e-4

(b) Caption of subfigure 2

Figure 5.5: Averaged correlations with linguistic features for the hearing condition.

		Classification accuracy (%)	
	Test Condition	Phones	Tones
Participant 1	Thinking	37.5	28.0
	Speaking	30.0	25.0
	Hearing	25.0	34.4
Participant 2	Thinking	35.0	26.0
	Speaking	37.5	40.6
	Hearing	18.8	28.1
	Average	30.6	30.3

Table 5.1: 4-class classification accuracies for 16 Chinese syllables. Chance classification in both "phones" and "tones" classification is 25 %.

Our results indicate that above-chance classification accuracy is possible in discriminating both segmental phonetic information (phone classification, as we have shown with the English data) and suprasegmental tonal information (Chinese phonetic tone). Averaged across our two speakers and three categories, accuracies for recognition of phones and tones were more or less identical (30.6 % vs 30.3 %). We can therefore suggest (only tentatively, due to the small sample size) that classification of phonetic tone for languages such as Chinese may be no more or less difficult than classification of phonemes.

5.4 Subject-independent results

For subject-independent testing, none of the models seemed to achieve above-chance classification accuracy for any of the three test conditions. Of the three, however, the SVM seems to achieve the best validation scores (See section 7.3 of the appendices).

5.5 Correlation of linguistic features with Electrodes

In order to find which electrodes were the source of the brain data features that were most strongly correlated with the linguistic features we wished to test, we used the same set of features that we calculated for our SVM. As described above, we calculated correlations for every feature by time window, resulting in a total of $19 \times 84 = 1596$ per electrode. We averaged over 1596×4 correlations-per-electrode in total to get the final results, since we were interested in getting average correlations across all 4 linguistic categorization tasks we tested (C/V, voiced/voicleass, $\pm/u/$, 4 vowels) in order to present a more balanced picture of where all linguistically-relevant information from our models is coming from.

To average over these features, we first converted the correlations to absolute correlations, since we are interested only in the strength of the correlation, and also different linguistic features could be positively or negatively correlated with the same brain-data

feature. We then calculated the z-score (using the arctanh function), took the average of the correlations, and then converted the z-score back into a Pearson correlation (r) using the tanh function. This method is described in more detail in [44]. The results are shown in figures 5.3-5.5.

For all three conditions, we see strong correlations with the FC5 electrode, which are particularly pronounced in the "stimuli" condition, where FC5 is the highest-correlated electrode. Two possible explanations relate to the fact that the FC5 electrode is close to Broca's area, linked to speech production (in fact, this is the closest area to the FC5 electrode on the average human skull [45]), or, alternatively, that it is reasonably close to the motor cortex, which is involved in the organization of motor movements in speech [46].

We also note that the visual cortex is positioned to the rear of the brain [47]. This explain the fact that correlations with the 01 electrode were so high (the second highest correlation). In order to test this in future experiments, we could provide just auditory stimuli instead of auditory and visual.

5.6 Correlation of linguistic features with EEG data features

The full list of correlations of linguistic features with EEG data features (divided into thinking, speaking, and hearing conditions) is shown in section 7.1 of the appendices.

Though there are differences between the three lists, it seems clear that in all three, temporal features are important, since out of the 10 most correlated features across the three conditions, all but 6 are delta or double-delta features.

Aside from this, we can see that measures of complexity appear to be key. For thinking data, the delta feature for Petrosian Fractal Dimensions is most highly correlated; for speaking data, the double-delta feature for Katz fractal dimensions is the third-most highly correlated, and for hearing data, the Katz fractal dimension feature is most highly-correlated.

These facts seem to show that the brainwaves associated with some phonetic categories are more complex than those associated with others. Discussing how these differences arise is beyond the scope of this investigation, but could be investigated in future work.

The delta or double-delta min + max features also appear in the top three most highly-correlated in each of the three conditions. These feature may help to show the overall trajectory and acceleration of the waveform; it is curious that this feature seems to be better than, for example, the delta mean or delta energy in discriminating this. These facts could also be explored in further research.

6. Conclusions

The conclusions that we make from this study are as follows:

- It is possible to use a lightweight (14 channel) EEG to obtain a dataset of comparable quality to one collected with a more standard 64 channel EEG (KARA-ONE) for use in imagined speech recognition.
- It is possible to use this dataset to obtain above-chance phone classification accuracy in both thinking, speaking, and hearing conditions.
- It is possible to use this dataset to distinguish a range of phonetic contrasts, including consonant vs. vowel, voiced vs. voiceless, $\pm$ /u/, four-class vowel recognition, and (Chinese) lexical tone.
- The most effective model that we managed to implement on our datasets was the SVM. However, judging from previous research, it is likely that CNN models could be effective on our data, but require further research to implement successfully.

6.1 FEIS vs KARA data

As indicated in section 5.1, differences between speaker-dependent classification accuracy using an SVM on our data and the KARA-ONE data was relatively small. Where accuracies on KARA data were better, this can be explained as the result of the fact that their categories had less internal variation (e.g. for their C/V classification task, there were only 5 consonants and 2 vowels, compared to our 8 consonants and 4 vowels). Where there were comparable levels of internal variation in both datasets (i.e. the $\pm$ /u/ task), there were no substantial differences between the FEIS and KARA data.

6.2 Thinking, speaking, and hearing conditions

As shown in section 5.2, above-chance classification accuracy was possible for all three test conditions: thinking (imagined speech), speaking, and hearing. Classification accuracy seemed to be better for thinking and hearing conditions and worse for the speaking condition, corroborating Zhao and Rudzicz's result in [3].

6.3 Comparison of linguistic features

Some linguistic features appeared to be easier to detect than others. From the three binary classes, the contrast with the highest classification accuracy was ± /u/, followed by consonant/vowel, followed by voiced/voiceless. These classification accuracies show an inverse correlation with the internal variation (number of phones) in the "smallest" class in each contrast: 1 in ± /u/, 4 in consonant/vowel, and 6 in voiced/voiceless.

Some interactions between linguistic category and test condition may be observed but require further research to show conclusively - for example, classification accuracy for voiced/voiceless appears to be substantially higher in the "hearing" condition compared to "thinking" or "speaking", possibly because voicing is most salient in this condition, being a linguistic feature with a very salient acoustic correlate.

6.3.1 Comparison of models

In all cases, the SVM obtained the best classification accuracy for our subject dependent data. For the speaker indpendent data, most models performed at a near-chance level, and so comparisons of the results are not especially useful.

Being a simpler model with a lower number of hyperparameters, the SVM faster to train than either of the other two models, and, relying on a large number of handcrafted features seems to make it fairly robust.

However, the SVM has some unattractive features compared to neural models, the most blatant of which is the long time required to calculate the large number of features (feature types × time windows × electrodes) prior to performing online decoding. In our implementation, it took on the order of half a minute, training on a powerful computing cluster, to calculate all the required features for 5 seconds of recording.

In this respect, the CNN model which took raw data as its input (the 2DCNN) might be most promising for future applications requiring real-time brain-computer interfacing. However, since this model performed around chance on most of our tasks, further research is needed to see if it can perform better with architectural changes or tuning of the model's hyperparameters.

The video classification style CNN (VCCNN) model had one test result which seemed to be above chance, but otherwise performed no better than the 2DCNN. While this model, like the SVM, requires computationally expensive pre-processing of the features to create an EEG "video", it has the advantage of creating a common representation of data from different devices which may make cross-device training easier, and for this reason warrants further research.

Bibliography

[1] B059691, "An investigation into the possibilities and limitations of decoding heard, imagined and spoken phonemes using the emotiv epoc+ mobile eeg headset," 2019.

[2] B. B059691, *Fourteen-channel EEG with Imagined Speech (FEIS) dataset*, Aug. 2019. DOI: 10.5281/zenodo.3369179. [Online]. Available: https://doi.org/10.5281/zenodo.3369179.

[3] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2015, pp. 992–996.

[4] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493, 2019.

[5] R. Srinivasan, "Methods to improve the spatial resolution of eeg," *International Journal of Bioelectromagnetism*, vol. 1, no. 1, pp. 102–111, 1999.

[6] C. Herff and T. Schultz, "Automatic speech recognition from neural signals: A focused review," *Frontiers in neuroscience*, vol. 10, p. 429, 2016.

[7] M. Lopez-Gordo, D. Sanchez-Morillo, and F. Valle, "Dry eeg electrodes," *Sensors*, vol. 14, no. 7, pp. 12 847–12 870, 2014.

[8] N. Birbaumer, "Breaking the silence: Brain–computer interfaces (bci) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.

[9] J. S. Brumberg, A. Nieto-Castanon, P. R. Kennedy, and F. H. Guenther, "Brain–computer interfaces for speech communication," *Speech communication*, vol. 52, no. 4, pp. 367–379, 2010.

[10] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[11] S.-G. Kim, W. Richter, and K. Ufffdfffdurbil, "Limitations of temporal resolution in functional mri," *Magnetic resonance in medicine*, vol. 37, no. 4, pp. 631–636, 1997.

[12] H. C. Sing, M. A. Kautz, D. R. Thorne, S. W. Hall, D. P. Redmond, D. E. Johnson, K. Warren, J. Bailey, and M. B. Russo, "High-frequency eeg as measure of cognitive function capacity: A preliminary report," *Aviation, space, and environmental medicine*, vol. 76, no. 7, pp. C114–C135, 2005.

[13] E. Estrada, H. Nazeran, P. Nava, K. Behbehani, J. Burk, and E. Lucas, "Eeg feature extraction for classification of sleep stages," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, vol. 1, 2004, pp. 196–199.

[14] T. Wang, J. Deng, and B. He, "Classifying eeg-based motor imagery tasks by means of time–frequency synthesized spatial patterns," *Clinical Neurophysiology*, vol. 115, no. 12, pp. 2744–2753, 2004.

[15] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, 2008.

[16] E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S. S. Meyer, L. D. Muñoz, K. J. Mullinger, T. M. Tierney, S. Bestmann, *et al.*, "Moving magnetoencephalography towards real-world applications with a wearable system," *Nature*, vol. 555, no. 7698, p. 657, 2018.

[17] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the p300 speller," *Journal of neural engineering*, vol. 3, no. 4, p. 299, 2006.

[18] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the emotiv epoc headset for p300-based applications," *Biomedical engineering online*, vol. 12, no. 1, p. 56, 2013.

[19] W. Speier, C. Arnold, J. Lu, R. K. Taira, and N. Pouratian, "Natural language processing with dynamic classification improves p300 speller accuracy and bit rate," *Journal of neural engineering*, vol. 9, no. 1, p. 016 004, 2011.

[20] M. A. Siegler and R. M. Stern, "On the effects of speech rate in large vocabulary speech recognition systems," in *1995 international conference on acoustics, speech, and signal processing*, IEEE, vol. 1, 1995, pp. 612–615.

[21] C. M. Colbert and D. Johnston, "Axonal action-potential initiation and na+ channel densities in the soma and axon initial segment of subicular pyramidal neurons," *Journal of Neuroscience*, vol. 16, no. 21, pp. 6676–6686, 1996.

[22] P. Olejniczak, "Neurophysiologic basis of eeg," *Journal of clinical neurophysiology*, vol. 23, no. 3, pp. 186–189, 2006.

[23] G. H. Klem, H. O. Lüders, H. Jasper, C. Elger, *et al.*, "The ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. 3, pp. 3–6, 1999.

[24] Emotiv, *Emotiv epoc*. [Online]. Available: `https://www.emotiv.com/product/emotiv-epoc-14-channel-mobile-eeg`.

[25] K. Kirihara, A. J. Rissling, N. R. Swerdlow, D. L. Braff, and G. A. Light, "Hierarchical organization of gamma and theta oscillatory dynamics in schizophrenia," *Biological psychiatry*, vol. 71, no. 10, pp. 873–880, 2012.

[26] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[27] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts: Comparison between ica and pca," in *Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No. 98TH8378)*, IEEE, 1998, pp. 63–72.

[28] G. D. Brown, S. Yamada, and T. J. Sejnowski, "Independent component analysis at the neural cocktail party," *Trends in neurosciences*, vol. 24, no. 1, pp. 54–63, 2001.

[29] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[30] B. N. Cuffin, "Effects of head shape on eegs and megs," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 1, pp. 44–52, 1990.

[31] K. Brigham and B. V. Kumar, "Subject identification from electroencephalogram (eeg) signals during imagined speech," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2010, pp. 1–8.

[32] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep learning the eeg manifold for phonological categorization from active thoughts," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 2762–2766.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[34] F. A. Heilmeyer, R. T. Schirrmeister, L. D. Fiederer, M. Volker, J. Behncke, and T. Ball, "A large-scale evaluation framework for eeg deep learning architectures," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2018, pp. 1039–1045.

[35] A. G. Lewis, J.-M. Schoffelen, H. Schriefers, and M. Bastiaansen, "A predictive coding perspective on beta oscillations during sentence-level language comprehension," *Frontiers in human neuroscience*, vol. 10, p. 85, 2016.

[36] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[37] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, "Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments," *Presence: teleoperators and virtual environments*, vol. 19, no. 1, pp. 35–53, 2010.

[38] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "Mne software for processing meg and eeg data," *Neuroimage*, vol. 86, pp. 446–460, 2014.

[39] Martinos, *Artifact correction with ica*. [Online]. Available: `https://martinos.org/mne/stable/auto_tutorials/preprocessing/plot_artifacts_correction_ica.html`.

[40] E. Başar, C. Başar-Eroglu, B. Rosen, and A. Schütt, "A new approach to endogenous event-related potentials in man: Relation between eeg and p300-wave," *International Journal of Neuroscience*, vol. 24, no. 1, pp. 1–21, 1984.

[41] N. McNair, *Entropy python package*. [Online]. Available: `https://pypi.org/project/EntroPy-Package/`.

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[43] R. T. Schirrmeister, *Braindecode python package*. [Online]. Available: `https://robintibor.github.io/braindecode/`.

[44] D. M. Corey, W. P. Dunlap, and M. J. Burke, "Averaging correlations: Expected values and bias in combined pearson rs and fisher's z transformations," *The Journal of general psychology*, vol. 125, no. 3, pp. 245–261, 1998.

[45] W. Klimesch, M. Doppelmayr, H. Wimmer, W. Gruber, D. Röhm, J. Schwaiger, and F. Hutzler, "Alpha and beta band power changes in normal and dyslexic children," *Clinical Neurophysiology*, vol. 112, no. 7, pp. 1186–1195, 2001.

[46] F. Pulvermüller, M. Huss, F. Kherif, F. M. del Prado Martin, O. Hauk, and Y. Shtyrov, "Motor cortex maps articulatory features of speech sounds," *Proceedings of the National Academy of Sciences*, vol. 103, no. 20, pp. 7865–7870, 2006.

[47] C. S. Herrmann, "Human eeg responses to 1–100 hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experimental brain research*, vol. 137, no. 3-4, pp. 346–353, 2001.

7. Appendices

7.1 Features used in SVM

This section shows a full list of all the "simple" (not including "delta" or "delta delta") features we calculated for use in the SVM.

7.2 Correlations with electrodes

Below is an averaged list of the absolute Pearson correlations of each feature we calculated for the SVM with the linguistic categories we examined. Each feature is an average of absolute correlations across 14 electrodes, 19 time windows and 4 linguistic features (C/V, voiced/voiceless $\pm$ /u/ and 4-class vowel categorization). Correlations were averaged according to the method described above - they were first converted to z-scores using the arctanh function, then the mean was taken, and then they were converted back to correlations using the tanh function.

7.3 Subject independent results

This section contains results obtained in the subject independent condition (training on all speakers but one, and testing on a single speaker). All results are % classification accuracies. We include results on our dataset (FEIS) the KARA-ONE dataset (KARA) and both datasets combined (KARA+FEIS). For the latter, we tested on speakers from the FEIS dataset.

Abbreviation	Full name
Mean	Mean average
Absmean	Mean of absolute values
Max	Maximum value
Absmax	Max of absolute values
Min	Minimum value
Absmin	Min of absolute values
App Entropy	Approximate entropy
Min+max	Maximum value + minimum value
Max-min	Maximum value - minimum value
Curvelength	Curve length
DFA	Detrended fluctuation analysis
EHF	Enhancement factor
Energy	Energy (sum of squared amplitudes)
Higuchi	Higuchi Fractal Dimensions
Integral	Simpson Integral
Katz	Katz Fractal Dimensions
Kurt	Kurtosis
NLE	Nonlinear energy
Perm entropy	Permutation Entropy
Petrosian	Petrosian Fractal Dimensions
Rootmeansquare	Root mean square
Skew	Skew
Samp entropy	Sample entropy
Spec entropy	Spectral entropy
Stddeviation	Standard deviation
SVD entropy	Spectral value decomposition entropy
Sum	Sum of all amplitudes
Variance	Variance

Table 7.1: A list of all features calculated from the EEG data for use in the SVM

Feature	Correlation (Pearson's r)	Feature	Correlation (Pearson's r)
dd petrosian	0.0115	d rootmeansquare	4.23e-3
dd minplusmax	8.72e-3	dd sum	4.17e-3
absmax	8.70e-3	d dfa	4.13e-3
d absmin	8.48e-3	dd nonlinear energy	4.11e-3
dd mean	7.92e-3	dd app entropy	4.07e-3
d app entropy	7.48e-3	integral	3.96e-3
kurtosis	7.35e-3	d mean	3.87e-3
maxminusmin	7.16e-3	perm entropy	3.84e-3
d absmax	6.92e-3	nonlinear energy	3.83e-3
d kurtosis	6.86e-3	d nonlinear energy	3.82e-3
d katz	6.83e-3	d energy	3.81e-3
sample entropy	6.69e-3	d svd entropy	3.77e-3
d maximum	6.67e-3	d integral	3.77e-3
absmean	6.54e-3	d perm entropy	3.75e-3
d minimum	6.53e-3	maximum	3.70e-3
minimum	6.34e-3	d petrosian	3.69e-3
d minplusmax	6.31e-3	d sample entropy	3.39e-3
dd energy	6.29e-3	katz	3.27e-3
dd minimum	6.26e-3	sum	3.23e-3
d maxminusmin	6.25e-3	curvelength	3.11e-3
d ehf	6.00e-3	dd dfa	2.98e-3
dd absmin	5.85e-3	app entropy	2.90e-3
energy	5.79e-3	variance	2.85e-3
d sum	5.71e-3	d absmean	2.83e-3
mean	5.56e-3	dd perm entropy	2.80e-3
dd higuchi	5.54e-3	dd maximum	2.69e-3
d stddeviation	5.51e-3	dd skew	2.63e-3
d curvelength	5.24e-3	d spec entropy	2.48e-3
ehf	5.16e-3	d skew	2.42e-3
dd spec entropy	5.12e-3	higuchi	2.37e-3
minplusmax	5.11e-3	skew	2.32e-3
svd entropy	5.05e-3	dd ehf	2.25e-3
dd absmean	4.90e-3	dd kurtosis	2.14e-3
dd katz	4.54e-3	d higuchi	1.89e-3
petrosian	4.49e-3	rootmeansquare	1.71e-3
spec entropy	4.46e-3	dd curvelength	1.64e-3
dd maxminusmin	4.44e-3	dd sample entropy	1.45e-3
dd svd entropy	4.38e-3	dd integral	1.44e-3
absmin	4.34e-3	stddeviation	1.42e-3
dd absmax	4.32e-3	dd variance	1.37e-3
dd rootmeansquare	4.31e-3	dd stddeviation	1.15e-3
dfa	4.25e-3	d variance	1.10e-3

Table 7.2: Correlations of 4 linguistic features with all mathematical features used (for thinking data)

Feature	Correlation (Pearson's r)	Feature	Correlation (Pearson's r)
d sum	0.0113	d kurtosis	4.94e-3
d minplusmax	8.62e-3	absmax	4.93e-3
dd katz	8.32e-3	d ehf	4.71e-3
dd sum	8.28e-3	dd integral	4.62e-3
dd petrosian	8.13e-3	dd svd entropy	4.57e-3
d dfa	8.10e-3	stddeviation	4.47e-3
d nonlinear energy	7.87e-3	d variance	4.43e-3
dd kurtosis	7.85e-3	petrosian	4.35e-3
higuchi	7.63e-3	svd entropy	4.34e-3
ehf	7.38e-3	d absmin	4.31e-3
maxminusmin	7.34e-3	curvelength	4.29e-3
dfa	7.17e-3	d sample entropy	4.29e-3
energy	6.94e-3	d petrosian	4.27e-3
dd spec entropy	6.94e-3	dd variance	4.24e-3
d katz	6.93e-3	d perm entropy	4.14e-3
sample entropy	6.80e-3	absmean	4.12e-3
dd nonlinear energy	6.70e-3	dd ehf	4.06e-3
dd dfa	6.60e-3	dd higuchi	4.05e-3
dd mean	6.45e-3	d skew	4.05e-3
dd minplusmax	6.26e-3	d mean	3.65e-3
absmin	6.16e-3	sum	3.65e-3
skew	6.13e-3	d spec entropy	3.44e-3
dd skew	6.08e-3	dd absmax	3.41e-3
perm entropy	6.04e-3	dd minimum	3.39e-3
rootmeansquare	6.01e-3	dd absmean	3.36e-3
integral	6.00e-3	d svd entropy	3.28e-3
mean	5.88e-3	dd absmin	3.26e-3
d maxminusmin	5.71e-3	d energy	3.25e-3
variance	5.70e-3	spec entropy	3.24e-3
dd curvelength	5.65e-3	d rootmeansquare	3.06e-3
dd stddeviation	5.60e-3	katz	3.06e-3
maximum	5.59e-3	d curvelength	3.01e-3
d stddeviation	5.56e-3	dd energy	2.93e-3
dd maximum	5.52e-3	d app entropy	2.91e-3
dd sample entropy	5.50e-3	dd rootmeansquare	2.69e-3
d higuchi	5.40e-3	dd maxminusmin	2.54e-3
app entropy	5.35e-3	d maximum	2.33e-3
minimum	5.29e-3	d integral	2.25e-3
d minimum	5.26e-3	kurtosis	2.11e-3
dd perm entropy	5.17e-3	dd app entropy	1.82e-3
d absmax	5.14e-3	minplusmax	1.64e-3
nonlinear energy	4.95e-3	d absmean	1.21e-3

Table 7.3: Correlations of 4 linguistic features with all mathematical features used (for speaking data)

Feature	Correlation (Pearson's r)	Feature	Correlation (Pearson's r)
katz	8.06e-3	dd katz	3.73e-3
dd sample entropy	7.79e-3	rootmeansquare	3.64e-3
d minplusmax	7.23e-3	kurtosis	3.56e-3
dd perm entropy	7.11e-3	skew	3.55e-3
dd spec entropy	6.55e-3	dd absmin	3.55e-3
d curvelength	6.35e-3	maximum	3.54e-3
dd curvelength	5.89e-3	sample entropy	3.40e-3
d dfa	5.77e-3	d higuchi	3.30e-3
dd higuchi	5.74e-3	absmean	3.30e-3
dd absmean	5.73e-3	higuchi	3.28e-3
minplusmax	5.71e-3	d katz	3.26e-3
d rootmeansquare	5.63e-3	energy	3.25e-3
spec entropy	5.53e-3	dd minplusmax	3.24e-3
d perm entropy	5.52e-3	dd absmax	3.22e-3
variance	5.46e-3	dd maxminusmin	3.18e-3
dd energy	5.46e-3	curvelength	3.15e-3
dd integral	5.44e-3	app entropy	3.12e-3
d kurtosis	5.21e-3	d spec entropy	3.11e-3
d app entropy	5.17e-3	d energy	3.05e-3
dd svd entropy	5.16e-3	d absmax	2.95e-3
minimum	5.14e-3	absmin	2.93e-3
integral	5.03e-3	d svd entropy	2.85e-3
svd entropy	5.02e-3	d absmin	2.84e-3
petrosian	5.01e-3	dd petrosian	2.83e-3
sum	5.01e-3	dd maximum	2.78e-3
d skew	4.97e-3	dd mean	2.65e-3
d nonlinear energy	4.76e-3	dd variance	2.52e-3
dd ehf	4.65e-3	dfa	2.49e-3
mean	4.55e-3	ehf	2.31e-3
dd kurtosis	4.53e-3	perm entropy	2.30e-3
dd dfa	4.51e-3	d maximum	2.23e-3
dd rootmeansquare	4.41e-3	d minimum	2.19e-3
d maxminusmin	4.38e-3	maxminusmin	2.19e-3
d ehf	4.15e-3	dd minimum	2.17e-3
dd nonlinear energy	4.14e-3	d variance	2.17e-3
d sum	4.04e-3	dd stddeviation	2.15e-3
d petrosian	3.98e-3	d integral	2.13e-3
d absmean	3.90e-3	d sample entropy	2.08e-3
nonlinear energy	3.89e-3	absmax	1.91e-3
stddeviation	3.85e-3	dd sum	1.88e-3
dd app entropy	3.84e-3	d stddeviation	1.60e-3
dd skew	3.80e-3	d mean	1.07e-3

Table 7.4: Correlations of 4 linguistic features with all mathematical features used (for hearing data)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	75.2	75.1	75.1	75.2	24.9	65.0	22.4
Hearing	75.2	25.0	75.1	58.4	75.0	61.8	21.8

Table 7.5: SVM (C/V): FEIS Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	46.3	50.0	50.0	46.3	50.0	48.52	2.0
Speaking	50.0	53.8	52.5	48.8	51.3	51.3	2.0
Hearing	50.0	50.0	50	51.3	51.3	50.5	0.7

Table 7.6: SVM (C/V): FEIS Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	67.6	75.2	75.2	24.8	75.7	63.7	22.0
Speaking	75.2	51.7	5.2	50.0	75.2	65.5	13.4
Hearing	75.2	25.0	75.1	58.4	75.1	61.8	21.7

Table 7.7: SVM (C/V): KARA Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	50.0	50.0	50.0	54.2	50.8	1.9
Speaking	54.2	52.1	50.0	50.0	43.8	50.0	3.9
Hearing	56.2	52.1	50.0	50.0	62.5	54.2	5.3

Table 7.8: SVM (C/V): KARA Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	50.0	75.1	75.1	75.1	75.2	70.1	11.2
Speaking	50.0	60.7	67.1	44.9	75.1	59.6	12.3
Hearing	25.0	75.5	75.5	50.5	55.8	56.4	20.9

Table 7.9: SVM (C/V): FEIS+KARA Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	50.0	51.3	50.0	50.0	50.3	0.6
Speaking	48.7	52.5	50.0	50.0	50.0	50.2	1.4
Hearing	50.0	48.8	51.3	51.3	50.0	50.3	1.1

Table 7.10: SVM (C/V): FEIS+KARA Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	50.0	50.2	55.3	53.8	52.0	52.3	2.3
Speaking	57.4	53.7	60.1	52.2	51.7	55.2	3.8
Hearing	25.0	75.2	62.6	50.0	50.0	52.6	18.6

Table 7.11: SVM (voiced/ voiceless): FEIS Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	47.5	45.0	45.9	50.0	47.5	47.2	1.9
Speaking	52.5	55.0	50.0	50.0	49.2	51.3	2.4
Hearing	50.8	47.5	49.2	49.2	50.8	49.5	1.4

Table 7.12: SVM (voiced/ voiceless): FEIS Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	62.9	62.5	62.9	64.0	55.3	61.6	3.5
Speaking	50.0	60.7	67.1	44.9	75.1	59.6	12.3
Hearing	56.8	54.5	66.3	62.0	54.5	58.8	5.2

Table 7.13: SVM (voiced/ voiceless): KARA Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	52.1	52.1	64.6	47.9	53.4	6.5
Speaking	50.0	54.2	50.0	54.2	58.3	53.3	3.5
Hearing	50.0	64.6	47.9	37.5	52.1	50.4	9.7

Table 7.14: SVM (voiced/ voiceless): KARA Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	55.0	75.2	75.1	52.1	53.4	62.2	11.9
Speaking	75.0	66.8	75.1	59.4	50.0	65.3	10.7
Hearing	75.1	56.3	58.4	55.0	59.1	60.8	8.2

Table 7.15: SVM (voiced/ voiceless): FEIS+KARA Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	50.0	49.2	50.0	50.0	50.0	49.8	0.4
Hearing	51.7	50.0	50.0	49.2	50	50.2	0.9

Table 7.16: SVM (voiced/ voiceless): FEIS+KARA Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	50.0	68.1	55.3	76.0	50.0	59.9	11.1
Speaking	50.0	63.2	58.7	41.3	75.3	57.6	12.9
Hearing	77.1	75.3	75.7	25.0	75.3	65.6	22.8

Table 7.17: SVM (± /u/): FEIS Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	45.0	45.0	50.0	50.0	50.0	48.0	2.8
Speaking	55.0	50.0	50.0	50.0	45.0	50.0	3.5
Hearing	50.0	50.0	45.0	50.0	50.0	49.0	2.2

Table 7.18: SVM (± /u/): FEIS Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	65.4	67.3	61.0	63.8	68.3	65.2	2.9
Speaking	75.4	76.4	68.6	75.4	41.2	67.5	15.0
Hearing	56.8	54.5	66.3	62.0	54.5	58.8	5.2

Table 7.19: SVM (± /u/): KARA Validation Scores (% Classification accuracy)

	Speaker							
	Validation	05	07	13	14	20	Average	Std. Dev
Thinking	66.4	58.3	62.5	50.0	50.0	45.8	53.3	6.8
Speaking	45.8	62.5	58.3	54.2	54.2	50.0	55.8	4.7
Hearing	50.0	64.6	47.9	37.5	52.1	50.0	50.4	9.7

Table 7.20: SVM (± /u/): KARA Test Scores (% Classification accuracy)

	Split						
	Split 0	Split 1	Split 2	Split 3	Split 4	Average	Std. Dev
Thinking	75.3	75.3	52.2	57.5	60.4	64.1	10.6
Speaking	75.2	58.5	55.2	75.3	75.7	68.0	10.2
Hearing	50.0	50.0	50.7	52.1	57.6	52.1	3.2

Table 7.21: SVM (± /u/): FEIS+KARA Validation Scores (% Classification accuracy)

	Speaker						
	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	45.0	50.0	50.0	50.0	50.0	49.0	2.2
Hearing	50.0	50.0	50.0	50.0	50.0	50.0	0.0

Table 7.22: SVM ($\pm$ /u/): FEIS+KARA Test Scores (% Classification accuracy)

	Speaker							
	Validation	05	07	13	14	20	Average	Std. Dev
Thinking	50.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	50.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Hearing	50.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0

Table 7.23: VCCNN (C/V): FEIS Test Scores (% Classification accuracy)

	Speaker							
	Validation	MM09	MM12	MM16	MM19	MM20	Average	Std. Dev
Thinking	50.5	50.5	50.5	50.5	50.5	50.5	50.5	0.0
Speaking	50.5	50.5	50.5	50.5	50.5	50.5	50.5	0.0
Hearing	50.5	50.5	50.5	50.5	50.5	50.5	50.5	0.0

Table 7.24: VCCNN (C/V): KARA Test Scores (% Classification accuracy)

	Speaker							
	Validation	05	07	13	14	20	Average	Std. Dev
Thinking	53.8	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	52.5	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Hearing	53.8	50.0	50.0	50.0	50.0	50.0	50.0	0.0

Table 7.25: VCCNN (C/V): FEIS+KARA Test Scores (% Classification accuracy)

	Speaker							
	Validation	05	07	13	14	20	Average	Std. Dev
Thinking	58.8	46.7	50.0	58.3	50.0	48.3	50.7	4.5
Speaking	66.3	56.3	53.8	45.00	48.8	47.5	50.3	4.6
Hearing	60.0	55.0	55.0	51.3	48.8	48.8	51.8	3.1

Table 7.26: 2DCNN (C/V): FEIS Test Scores (% Classification accuracy)

	Speaker							
	Validation	05	07	13	14	20	Average	Std. Dev
Thinking	55.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	50.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Hearing	50.5	50.0	50.0	50.0	50.0	50.0	50.0	0.0

Table 7.27: 2DCNN (C/V):KARA Test Scores (% Classification accuracy)

	Speaker							
	Validation	MM09	MM12	MM16	MM19	MM20	Average	Std. Dev
Thinking	50.5	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Speaking	55.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
Hearing	50.5	50.0	50.0	50.0	50.0	50.0	50.0	0.0

Table 7.28: 2D CNN (C/V): FEIS+KARA Test Scores (% Classification accuracy)