On the Relative Contribution of Deep Convolutional Neural Networks for SSVEP-based Bio-Signal Decoding in BCI Speller Applications

Joshua J. Podmore [1], Toby P. Breckon [2, 3], Nik K. N. Aznan [2, 3] & Jason D. Connolly [1]
Department of {Psychology[1] | Computer Science[2] | Engineering[3]}, Durham University, Durham, U.K.

***Abstract*—Brain-computer interfaces (BCI) harnessing Steady State Visual Evoked Potentials (SSVEP) manipulate the frequency and phase of visual stimuli to generate predictable oscillations in neural activity. For BCI spellers, oscillations are matched with alphanumeric characters allowing users to select target numbers and letters. Advances in BCI spellers can, in part, be accredited to subject-specific optimization, including; 1) custom electrode arrangements, 2) filter sub-band assessments and 3) stimulus parameter tuning. Here we apply deep convolutional neural networks (DCNN) demonstrating cross-subject functionality for the classification of frequency and phase encoded SSVEP. Electroencephalogram (EEG) data are collected and classified using the same parameters across subjects. Subjects fixate forty randomly cued flickering characters (5 × 8 keyboard array) during concurrent wet-EEG acquisition. These data are provided by an open source SSVEP dataset. Our proposed DCNN, PodNet, achieves 86% and 77% offline Accuracy of Classification across-subjects for two data capture periods, respectively, 6-seconds (information transfer rate=40bpm) and 2-seconds (information transfer rate= 101bpm). Subjects demonstrating sub-optimal (<70%) performance are classified to similar levels after a short subject-specific training period. PodNet outperforms filter-bank canonical correlation analysis (FBCCA) for a low volume (3-channel) clinically feasible occipital electrode configuration. The networks defined in this study achieve functional performance for the largest number of SSVEP classes decoded via DCNN to date. Our results demonstrate PodNet achieves cross-subject, calibrationless classification and adaptability to sub-optimal subject data and low-volume EEG electrode arrangements.**

***Index Terms*—Brain-Computer Interface, Deep Convolutional Neural Network, Electroencephalography, Steady State Visual Evoked Potential, Calibrationless, BCI, DCNN, CNN, EEG, SSVEP.**

I. Introduction

BRAIN-computer interfaces (BCI) are integrated hardware and software ensembles that control assistive devices with brain-based bio-signals [1]. Electroencephalograms (EEG) are widely used for BCI research due to high temporal resolution, short set-up times and a low risk, non-invasive acquisition process [2]. The most effective EEG-based BCI to date [3] feature substantial optimization of data acquisition tools and stimuli presentation methods at the single-subject level. The move toward subject-specific analyses is due to individual differences in EEG data as well as the non-stationary and non-linear characteristics of EEG. Researchers have developed offline methods to find ideal EEG electrode configurations [4], [5], subject-specific frequency bands with Common Spatial Pattern algorithms [6] and stimulus parameter tuning [3]. Online BCI calibration has also been explored to improve signal-to-noise ratio using adaptive EEG reference selection [7].

In this study, we show calibrationless BCI decoding for the highest number of EEG-based Steady State Visual Evoked Potential (SSVEP) classes (40) with deep convolutional neural networks (DCNN) to date. SSVEP are frequency and phase dependent bio-signals detected from EEG electrodes over occipital and occipital-parietal regions [8]. These bio-signals are used in the development of BCI communication devices, known as spellers, which aim to provide users with the ability to converse with caregivers, clinicians and family members [2].

In BCI speller paradigms, visual stimuli typically consist of a target matrix containing alphanumeric characters and grammatical symbols presented via a computer monitor. A stimulus square is positioned behind each target and assigned a unique stimulus flicker pattern varying in frequency (Hz) and or phase. A stimulus square coded with an 8Hz flicker rate and zero-degree phase shift switches between white and black at intervals of 125ms (1000ms/8Hz). Fixation of a target paired to a stimulus square with these same properties leads to the propagation of an SSVEP waveform with the same oscillatory characteristics. Stimulus patterns embedded in the EEG signal timecourse are extracted and analyzed to determine the target intended for spelling.

SSVEP latency is low (150-350ms) and less prone to degradation over time, as compared to alternative bio-signals [9]. This makes SSVEP ideal for attaining high information transfer rates (ITR) during long-term real-world use. The ITR statistic defines BCI communication speed, expressed in bits per minute (bpm). This indicates device usability as ITR is dependent on; 1) Accuracy of Classification (AoC), 2) the number of stimulus targets; and 3) the data capture duration per target (refer to equation in section 3.2).

In this study, we apply a DCNN, PodNet, to extract stimulus pattern features embedded in EEG data for the classification of SSVEP target stimuli. DCNN use multiple convolutional layers to extract data features necessary to differentiate between classes. In this work, classes refer to the unique stimulus patterns assigned to stimulus squares. Convolutions involve iterative matrix multiplications across input data as confined by sliding windows with adjacent filter matrices. This leads to the extraction of data features relevant to the task of classification into feature maps [10]. We test the performance of our proposed DCNN, PodNet, in the offline classification of wet-EEG SSVEP-based bio-signals differing in both frequency and phase for a 40-class problem.

The networks are trained and tested for two data capture durations per target, 6 and 2 seconds. This is done to evaluate network performance at low data capture durations, as in real-world online use this leads to higher ITR. Network performance is also evaluated using data in which redundant samples captured immediately before or after the SSVEP stimulus period are excluded. In clinical settings, EEG-based communication aids must balance the number of electrodes used with functional performance. EEG devices using a lower number of electrodes ultimately translate to lower; 1) hardware costs, 2) hygiene risk, and 3) user discomfort [11]. The analyses are then tested using alternative electrode arrangements ranging from 10 to 3-channels.

PodNet performance is compared against a shallow CNN architecture, 1D SSVEP Convolutional Unit (1DSCU) [12] and Filter-Bank Canonical Correlation Analysis (FBCCA) [3]. We perform a single-subject level analysis and find PodNet performs sub-optimally ($<$70%) for two subjects (29 & 33). Individual pretrained PodNets are re-trained at the single-subject level to test network adaptability. All data used in network training, validation and testing are derived from a benchmark dataset provided by [13]. The PodNet DCNN defined herein demonstrate cross-subject classification of wet-EEG based bio-signals for novel, naïve BCI subjects and show adaptation in response to sub-optimal performance after a short subject-specific training period. Further, these networks outperform FBCCA for a 3-channel (occipital) electrode arrangement, making these techniques more viable for long-term use in clinical settings. Moreover, the significant contribution of the study is that PodNet achieves $>$85% cross-subject accuracy for the highest number of BCI-based bio-signal target classes (40) using DCNN [14], [15], [16].

II. Related Work

DCNN applications in EEG-based BCI research include mental imagery decoding, primarily for mobility applications [17], [18] and Event-Related Potential (ERP) detection, primarily for BCI speller applications [19], [20]. The first study to use DCNN in SSVEP classification [14] is implemented for a 5 class problem. The network uses two convolutional layers to develop spatial and time filters, with an embedded fast-Fourier transform to amplify frequency representations in feature maps. An AoC of 95.61% is achieved at the single-subject level. [15] shows a DCNN, EEGNet, consisting of 4 convolutional layers decodes four brain-based bio-signals; 1) ERP, 2) movement-related cortical potentials, 3) sensory-motor rhythms, and 4) error-related negativity responses. EEGNet matches or exceeds each of the leading classification methods for respective bio-signals both within-and across-subjects. This demonstrates DCNN can extract a wide range of BCI bio-signals.

In [16], a DCNN classifies 5 SSVEP targets providing directional instructions for an exoskeleton. CNN significantly outperforms canonical correlation analysis (CCA) and a hybrid CCA and k nearest neighbours method in the real-time classification of SSVEP during both static (99.28%) and ambulatory (92.28%) operation. This demonstrates that CNN can accommodate for EEG movement artefacts.

In [21], CNN trained using cross-subject and single-subject data are compared for the classification of Rapid Serial Visual Presentation (RSVP) EEG-based bio-signals. The larger cross-subject datasets for network training enhance the quality of feature maps, resulting in higher classification accuracy ($>$90%). Along these very same lines, [12] implement cross-subject training for the 1DSCU CNN in the classification of 4 unique SSVEP as collected using a Dry-EEG system. The network contains one convolutional layer and outperforms Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Minimum Distance to Mean (MDM) and Recurrent Neural Networks (RNN). Cross-subject training is used in the optimization of all PodNet configurations defined in this study and the 1DSCU is implemented as a baseline assessment.

Currently, FBCCA is the gold-standard in the classification of frequency and phase encoded SSVEP and consistently achieves $>$90% cross-subject accuracy [22], [3]. This technique involves performing CCA across original EEG data and filter transformed data to expose fundamental and harmonic SSVEP frequency components. FBCCA is also used as a baseline assessment to evaluate the performance of PodNet. In contrast to earlier CNN-based approaches, we implement PodNet for the classification of both frequency and phase embedded SSVEP. PodNet classifies the highest number of SSVEP targets (40) using DCNN to date [14], [15], [16]. Moreover, the network proposed herein is capable of $>$90% calibrationless classification on 95% of subjects. PodNet outperforms baseline assessments for a clinically feasible electrode arrangement and demonstrates the ability to extract task-relevant SSVEP data features from data to which networks initially perform sub-optimally.

III. Methodology

In this section, we discuss the experimental setup implemented in the collection of SSVEP via wet-EEG. The datasets used to evaluate both PodNet and baseline assessments are outlined. The methods used to provide baseline assessments, 1DSCU and FBCCA, are explored. Additionally, the proposed PodNet architecture is defined. We detail the methods used for evaluating the analyses on alternative electrode arrangements. Finally, we describe how to perform single-subject optimization with pretrained PodNets for two subjects (29 & 33) demonstrating sub-optimal performance.

A. Experimental Setup

We present subjects with a 5×8 visual target array of English alphabet characters (26), numbers (10) and symbols (4). A 60Hz monitor displays the targets centrally oriented inside stimulus squares, with a uniform distance between adjacent targets. Stimulus presentation code is programmed using MATLAB and Psychophysics Toolbox Release 2017a (MathWorks, Inc., Natick, USA). Subjects are cued to fixate targets (5s) via the presentation of a red square overlaid onto the target (0.50s), after which the monitor returns to a black screen resting stage (0.50s). One block consists of fixating all 40 targets once. A total of 6 blocks are collected for each subject. Wet-EEG samples are collected at a rate of 1000Hz. For the purposes of this study, analyses are restricted to the

following 10 electrodes, PO8, PO7, PO6, PO5, PO4, PO3, POz, O2, O1, and Oz.

40 unique target stimulus patterns are used following the approximation method in which joint frequency and phase values are assigned to each stimulus square [23]. Stimulus frequencies are limited by the optimal range for SSVEP propagation (8-15Hz) and monitor refresh rates. A 60Hz monitor can display stable stimulus patterns at 8.57Hz (7 frames per second), 10Hz (6 fps), 12Hz (5 fps) and 15Hz (4 fps), as when the refresh rate is divided by these values it produces an integer-valued frame-rate. Non-integer frame-rates are achieved by interleaving two stimulus frequencies within a 60Hz cycle [24], [22]. This increases the number of possible stimuli patterns within the optimal SSVEP range and allows for the implementation of a dense 40-target visual array. Each target is assigned a unique stimulus frequency ranging from 8.0Hz-15.8Hz, in 0.2Hz increments. Neighbouring frequencies are assigned phase values differing by 0.5π .

B. Datasets for Classification

The open source SSVEP data used in the training and testing of all networks defined in this study is provided by [13]. Data are collected from 35 subjects. A subset (n = 8) of the subjects have experience with SSVEP-based BCI spellers. The remaining subjects (n = 27) have no prior experience with BCI systems and are classed as BCI-naïve (see [13] for further details). The network test and validation datasets this study contain only, novel BCI-naïve subjects to simulate real-world usage. Ethics approval was provided by the relevant institutional committee [13]. Data preprocessing involves downsampling data from 1000Hz to 250Hz and normalization between -1 and +1.

Networks are evaluated on a range of datasets differing in the duration of data capture per target, stimulus exclusive data (refer to Table I) and electrode arrangements (refer to Table II). In all instances, the datasets are parsed into training, validation and test sets. Of the 35 subjects, 25 are used in cross-subject network training, 5 are used as a validation set to monitor network overfitting and 5 are used as a test set. The performance metrics reported here are calculated exclusively for 5 novel BCI-naïve subjects in the test dataset.

Datasets	Samples	Selection Time (Seconds)
6Sec	1500	6
2Sec	500	2
5Sec	1250	5.5
1.5Sec	375	2

TABLE I: Datasets for network evaluation are presented specifying, dataset naming scheme (Datasets), the number of samples per trial (Samples) and the length of the data capture period per target (Selection Time). These selection times are used in ITR calculations, (refer to equation below). The 6Sec dataset includes samples collected over the entire 6 Second trial period. The 2Sec dataset includes samples from the first 2 seconds. The 5Sec dataset excludes samples from the first 500ms of the trial period and samples collected in the final 500ms resting stage. The 1.5Sec dataset includes samples collected for 1.5 seconds after the initial 500ms visual search period.

This is done to simulate real-world BCI speller applications and ensure no subject data used during weight optimization is included in network evaluation. The purpose of training the network on different data capture sizes is to test network capabilities on low quantities of data, as shorter capture windows are crucial to enhancing system performance in real-world settings. Differences in the inclusion and exclusion of redundant samples in the data are tested primarily to reveal any potential benefits for DCNN.

To explore the applicability of analyses in clinical settings, performance is evaluated when restricting the number of electrodes sampled. Analyses using a high number of electrodes is not feasible for long-term clinical use due to; 1) patient discomfort, 2) device hygiene maintenance and 3) electrode replacement costs. Four datasets are used featuring alternative electrode arrangements. As shown in Table II, the electrodes sampled incrementally regress towards locations positioned over the occipital lobe. This is based on previous source localization studies demonstrating that the SSVEP waveform propagates primarily from the occipital lobe [25]. All data used is collected exclusively from the first 2 seconds of the trial period (refer to Table II).

Datasets	N	Electrodes
10Chan	10	O1, Oz, O2, POz, PO3, PO4, PO5, PO6, PO7, PO8
8Chan	8	O1, Oz, O2, POz, PO3, PO4, PO5, PO6
6Chan	6	O1, Oz, O2, POz, PO3, PO4
3Chan	3	O1, Oz, O2

TABLE II: Datasets for network evaluation are presented specifying the dataset naming scheme (Datasets), the number of channels per arrangement (N), and the electrode locations of each arrangement. These datasets are compiled from the 2Sec and 1.5Sec datasets (refer to Table I). These four electrode arrangements are chosen at each step from the 10Chan to 3Chan electrode arrays, restricting the brain surface sampled as a function of distance from the inion (cranial protuberance indicating occipital lobe positioning).

PodNets are evaluated using AoC, Average Precision (AP) and ITR. AoC refers to the number of correctly identified targets users intended to spell relative to the total number of targets users are instructed to fixate. AP refers to single class selection, relative to all other classes. As defined in [1], in computing ITR, N refers to the number of potential targets for classification. P denotes the probability of target classification. T indicates selection time duration per trial (refer to Table I).

$$B = log_2 N + P\, log_2\, P + (1-P)\, log_2[(1-P)/(N-1)]$$

$$Bits\, per\, minute = B \times (60\,/\,T)$$

C. Baseline Assessments

To evaluate PodNet performance we compare against two baseline assessments. Based on the recent results of [12] the 1DSCU is evaluated to demonstrate the effectiveness of shallow CNN for the task of classifying frequency and phase encoded SSVEP. The 1DSCU is constructed from an operational unit containing a one-dimensional convolutional, batch normalization and max-pooling layer.

FBCCA is applied using the open-source MATLAB toolkit defined in [3]. EEG signal timecourse data are transformed

via bandpass filters exposing harmonic frequency components of embedded SSVEP. Each filter-bank is processed using CCA to determine the frequency and phase of the SSVEP present within the original EEG data. Associated single-subject calibration procedures typically include fine-tuning of stimulus durations and phase offsets [3], [13]. Additionally, CCA reference signals are often blended with subject-specific EEG data [26], [3]. As this study aims to test calibrationless functionality, these procedures are not implemented. Of note, FBCCA is evaluated only in the analyses of datasets in which redundant data is removed (5sec and 1.5Sec) as this is a standardized practice when implementing the technique [22], [3].

D. Deep Convolutional Neural Network Design

All PodNet variants are implemented using the MatConvNet convolutional neural network library [27]. See Fig. 1 for the PodNet configuration applied in the classification of the 6Sec dataset. Each PodNet is configured with consistent architectural features. Operational layers are clustered into Pods consisting of convolutional, drop-out (50%), batch normalization, Rectified Linear Unit, and max pooling layers. Additionally, the last Pod contains a final dense layer which outputs to a softmax operation. This arrangement is consistent throughout PodNet configurations. All network weights are initialized using the Xavier method [28] and updated following the Adam optimization algorithm [29].

In the first Pod, 10×30 convolutional filters are used. This filter width (30) is used to capture the entire phase cycle of the lowest target class stimuli frequency (8.0Hz). Assigning filter depth to the number of channels sampled (30) is a convenient way of performing the well-established EEG research practice of average referencing. In this context, matrix multiplications are performed across both time and channel dimensions producing a 1D output array. All subsequent convolution filters are one-dimensional of size: 1×30. 2D Max-pooling operations of size: 2×2, with stride: 2 and pad: 1 used throughout.

PodNet configurations implemented for the classification of datasets 2Sec and 1.5Sec are constructed from 4 of the above-mentioned Pods. This is due to the reduced size of input data and our aim to ensure consistent convolution filter dimensions (size: 1×30). This leads to a lower number of Pods, with greater convolution and max pooling stride values. These PodNet configurations feature the same level of drop-out (50%) across convolutions, excluding the final convolutional layer in all instances. Each network is optimised over 2000 training epochs. Meta-parameters are consistent across networks in terms of objective function (stochastic gradient descent), learning rate (1e-5), momentum (0.9), and batch size (5).

E. Single-Subject Optimization

Typically, researchers train CNN at the single-subject level for SSVEP-based BCI speller applications [14]. This is not viable for the deep networks as the degree of data collection we require for effective weight optimization would constitute an inordinate amount of testing time for an individual subject. We implement a viable method of single-subject deep net optimization by using a PodNet, pretrained using data from multiple subjects. This is then re-trained using data exclusively from one subject.

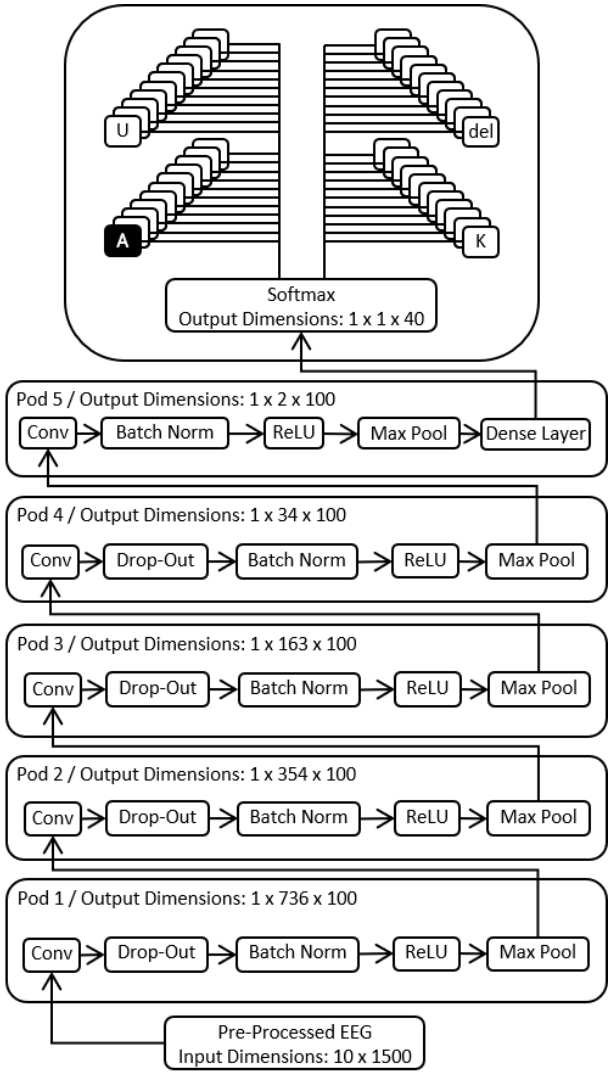

Fig 1. Diagram of PodNet configuration implemented for the 6Sec dataset, refer to Table I. Each Pod contains convolution (Conv), drop-out (Drop-Out), batch normalization (Batch Norm), Rectified Linear Unit (ReLU) and max pooling (Max Pool) layers. In the test configuration, batch normalization and drop-out layers are by-passed. Output dimensions refer to the depth and width of the data after passing through the final layer of each Pod.

To evaluate single-subject optimization using PodNet, two subjects (29 & 33) for which PodNet demonstrates sub-optimal performance are selected. This involves using a PodNet pretrained on the 6Sec dataset and re-training this network at the single-subject level. Each subject provides 6 blocks of EEG data (refer to Methods, Section A). Four of these blocks are used for PodNet re-training, one block is used as a validation set and one block is utilized for network evaluation. Single-subject PodNets are trained for a total of 500 epochs, with batch size: 2, due to low data volume. All additional meta-parameters as detailed above are maintained.

IV. Results and Discussion

A. Data Capture Comparison 6 vs 2 Seconds

The results for the classification of datasets 6Sec and 2Sec are presented in Table III. PodNet substantially outperforms the 1DSCU network in terms of AoC and ITR for both the 6Sec and 2Sec datasets. The high number of classes (40) makes the use of confusion matrices unfeasible. PodNet cross-class performance is presented using a graphical format in which AoC is plotted against AP. As shown in Fig. 2, PodNet demonstrates minimal variation in cross-class AoC and AP.

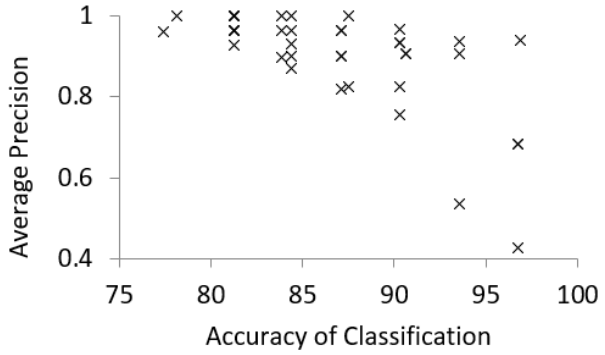

Fig 2. Plotting within class AoC and within class AP of PodNet for the 6Sec dataset. This contains 6 seconds of data capture per target. Each data point represents one of the 40 stimulus pattern classes differing in terms of both frequency and phase (refer to Table I).

PodNet shows substantially reduced performance and higher cross-class variation for the 2Sec dataset (refer to Fig. 3). This is due to the substantial reduction in data capture durations per target. BCI-spellers using this PodNet configuration would have an increased need for spelling correction and a higher speed of operation given the substantially higher ITR (refer to Table III).

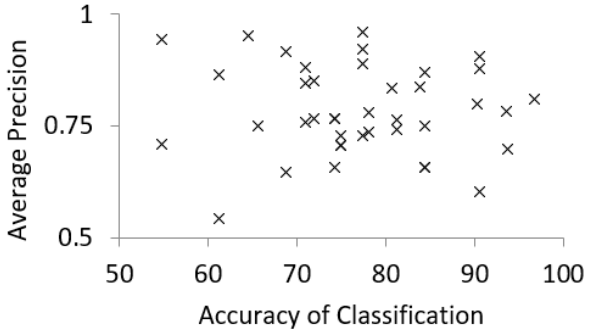

Fig 3. Plotting within class AoC against within class AP of PodNet for the 2Sec dataset. This contains 2 seconds of data capture per target including samples from the initial 500ms visual search trial period (refer to Table I).

Datasets	Method	Samples	Selection Time	AoC	ITR
6Sec	PodNet	1500	6	**86.75**	**40.57**
2Sec	PodNet	500	5.5	77.06	99.98
6Sec	1DSCU	1500	2	**74.73**	**31.71**
2Sec	1DSCU	500	2	39.49	11.56

TABLE III: Cross-Subject AoC and ITR for PodNet and the 1DSCU network evaluated using the 6Sec and 2Sec datasets.

B. Stimulus Exclusive Data Evaluation

Classification results of PodNet and baseline assessments for the 5Sec and 1.5Sec datasets are shown in Table IV. The removal of redundant data for the 5Sec dataset leads to no substantial changes in PodNet AoC or ITR. As shown in Fig. 4, the variance in decode performance does not change substantially for the PodNet optimized using the 5Sec dataset. The 1DSCU demonstrates substantially reduced performance for the 5Sec dataset as compared to the 6Sec dataset.

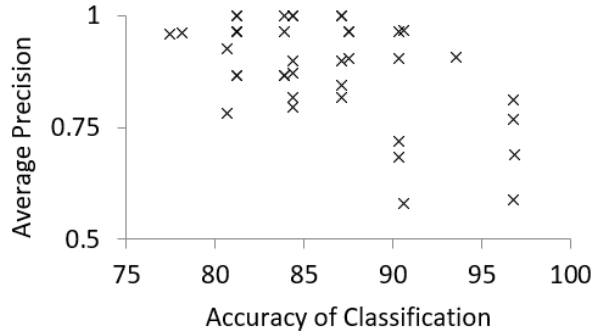

Fig 4. Plotting within class AoC and AP for PodNet, trained and tested using the 5Sec dataset. This contains 5 seconds of data capture per target, with redundant, non-stimulus data excluded (refer to Table I).

PodNet performance decreases marginally for the 1.5Sec dataset with pre-stimulus visual search data (500ms) removed. This is due to redundant data representing 25% of the total samples per trial. Within-class performance, as compared to PodNet optimized for the 2Sec dataset, shows marginally higher variance in AoC and AP (see Fig. 3 & 5).

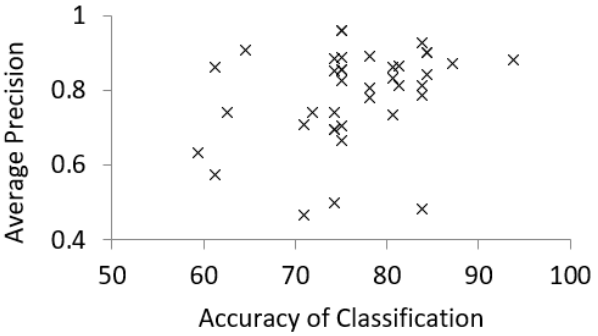

Fig 5. Plotting within class AoC and AP for PodNet, trained and tested using the 1.5Sec dataset. This excludes samples collected during the initial 500ms pre-stimulus visual search period.

FBCCA outperforms both PodNet and 1DSCU for the 5Sec and 1.5Sec datasets. PodNet utilizes four more convolutional layers than the 1DSCU. This added computation leads to the development of richer feature maps for the differentiation of classes, most markedly in low data capture contexts. The enhanced performance suggests increasing network depth can substantially improve AoC and ITR. It is suggested that an increase in the availability of large open source datasets would assist in the development of deeper networks with greater classification performance. The collection of EEG data at higher sampling rates (>250Hz) would increase the likelihood of networks developing dense SSVEP feature representations. Although speculative, it is predicted that deeper PodNet architectures, with a larger, higher resolution training dataset would show significantly higher AoC and ITR.

Datasets	Method	Samples	Selection Time	AoC	ITR
5Sec	PodNet	1250	5.5	**86.19**	**43.78**
5Sec	1DSCU	1250	5.5	68.63	30.18
5Sec	FBCCA	1250	5.5	97.92	50.66
1.5Sec	PodNet	500	2	**75.64**	**96.99**
1.5Sec	1DSCU	500	2	32.67	9.29
1.5Sec	FBCCA	500	2	84.00	115.25

TABLE IV: Cross-Subject accuracy of classification and ITR for the PodNet, 1DSCU and FBCCA evaluated using the 5Sec and 1.5Sec datasets

C. Electrode Arrangement Assessment

Analyses performance is evaluated using 4 dataset subsets comprising 4 unique electrode arrangements (refer to Table II). Previous source localization research suggests the locus of SSVEP signal propagation is positioned over the occipital lobes, with signal degradation occur as a function of the distance from the inion (external occipital protuberance). The electrodes comprising the 4 subsets are stepped from 10 to 3-channels in order to restrict samples collected as a function of distance from this biomarker.

All analyses are tested using only the first two seconds of data sampled, as this data collection parameter is more functional in real-world settings as compared to the 6-second data capture duration of the 6Sec dataset. PodNet is trained and tested using the 2Sec dataset, FBCCA is tested using the 1.5Sec dataset (see, Baseline Assessments). The 1DSCU is not evaluated as previous assessments below the 6-second data capture duration show performance below functional use (refer to Table III and IV).

	FBCCA		PodNet	
Datasets	AoC	ITR	AoC	ITR
10Chan	84.00	115.258	77.06	99.98
8Chan	84.33	116.028	77.78	101.50
6Chan	81.25	109.041	72.54	90.68
3Chan	62.083	70.8129	**66.67**	**79.26**

TABLE V: Cross-Subject accuracy of classification and ITR for the PodNet and FBCCA networks evaluated using the 6Sec and 2Sec datasets.

All analyses drop substantially in terms of AoC and ITR in response to a reduction of input electrodes (refer to Table V). For higher-volume electrode arrangements, FBCCA shows the highest AoC and ITR. FBCCA also demonstrates the largest drop in performance when comparing 10 to 3-channel dataset evaluation. PodNet performance shows higher resilience to the removal of electrodes, with a greater AoC and ITR as compared to FBCCA for the 3-channel dataset. This suggests PodNet has greater potential for long-term use in clinical settings.

D. Single-Subject Optimized PodNet Classification

The pretrained PodNet used for single-subject optimization is trained using the 6Sec dataset with both subject 29 and 33 removed. Cross-subject results for this PodNet configuration include an AoC of 98.34% and ITR of 51.12bpm. Prior to single-subject optimization, PodNet achieves an AoC of 75.00% for subject 29 and an ITR of 31.89bpm. Post-single-subject training, PodNet AoC increases to 90.00% and ITR increases to 43.24bpm (refer, to Fig. 6).

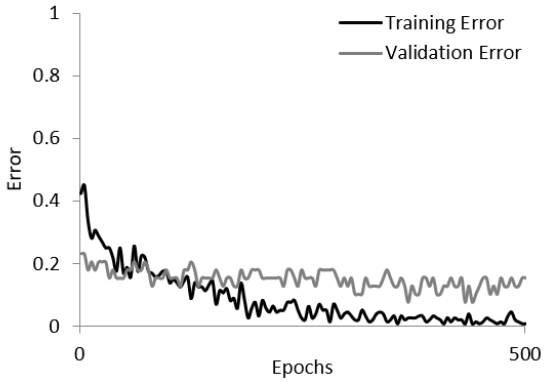

Fig 6. The graph shows the error loss (inverse of AoC) of PodNet on training and validation data subsets for subject 29 data over 500 epochs.

For subject 33 the pretrained PodNet achieves an AoC of 35.00% and ITR of 10.38bpm. Post-single-subject optimization, PodNet AoC increases to 70.00% and ITR increases to 28.55bpm, bringing performance up to the threshold of functional use (refer to Fig. 7).

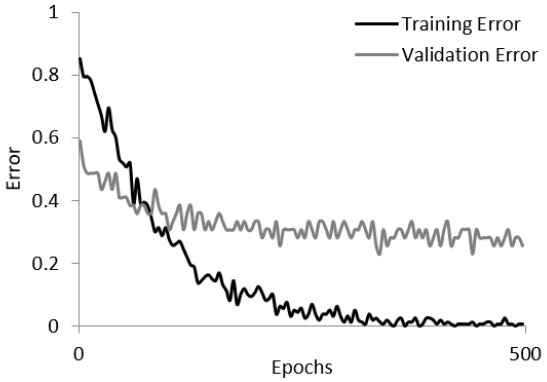

Fig 7. The graph shows the error loss (inverse of AoC) of PodNet on training and validation data subsets for subject 33 data over 500 epochs.

From the results presented (Table III, IV & V), we demonstrate that the performance of the proposed PodNet substantially outperforms prior DCNN-based work [14], [15], [16], [12] in terms of AoC and ITR, relative to the number of SSVEP targets decoded (40). PodNet evaluated using the 6Sec dataset achieves a mean AoC of 86.75% and ITR of 40.57bpm, and for the 1.5Sec dataset achieves a mean AoC of 75.64% and an ITR of 96.99bpm, for offline classification. Furthermore, the broad scope of our approach encompasses 40 joint frequency and phase encoded classes and is capable of SSVEP classification with low-volume electrode configurations and substantially enhances initially sub-optimal network performance after a short subject-specific optimization period, all of which are firsts for DCNN decoding of SSVEP.

V. Conclusion

The present study developed a DCNN, PodNet, for frequency and phase encoded SSVEP classification to address the absence of deep networks for high target density (>5 classes) BCI-based bio-signal decoding. The DCNN, as developed, outperform and extend the prior DCNN-based work of [14], [15], [16], [12]. Crucially, for 95% of subjects, >90% AoC is demonstrated in the absence of any subject-specific optimization. For the first time, DCNN are optimized at the single-subject level using pretrained networks to enhance sub-optimal network performance. Crucially, PodNet outperforms FBCCA, currently the gold-standard in SSVEP decoding, for a clinically feasible occipital electrode arrangement. The development of robust analytical methods benefits BCI users allowing for faster, accurate classification and long-term clinical use.

Acknowledgement

The authors would like to thank members of the Department of Biomedical Engineering Tsinghua University, Beijing, for access to their SSVEP EEG dataset [13].

References

[1] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 326–333, 1998.

[2] D. McFarland and J. Wolpaw, "EEG-based brain–computer interfaces," *Current Opinion in Biomedical Engineering*, vol. 4, pp. 194–200, 2017. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S246845111730082X

[3] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proceedings of the National Academy of Sciences*, vol. 112, no. 44, pp. E6058–E6067, 2015. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1508080112

[4] H. J. Hwang, J. H. Lim, J. H. Lee, and C. H. Im, "Implementation of a mental spelling system based on steady-state visual evoked potential (SSVEP)," *2013 International Winter Workshop on Brain-Computer Interface, BCI 2013*, pp. 81–83, 2013.

[5] J. H. Lim, J. H. Lee, H. J. Hwang, D. H. Kim, and C. H. Im, "Development of a hybrid mental spelling system combining SSVEP-based brain-computer interface and webcam-based eye tracking," *Biomedical Signal Processing and Control*, vol. 21, pp. 99–104, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2015.05.012

[6] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, 2012.

[7] Z. Wu and S. Su, "A dynamic selection method for reference electrode in SSVEP-based BCI," *PLoS ONE*, vol. 9, no. 8, 2014.

[8] A. M. Norcia, L. G. Appelbaum, J. J. M. Ales, B. B. R. Cottereau, and B. Rossion, "The steady-state visual evoked potential in vision research: a review," *Journal of Vision*, vol. 15, no. 6, p. 4, 2015. [Online]. Available: http://jov.arvojournals.org/article.aspx?articleid=2291652

[9] F. B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives," pp. 418–438, 2010.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[11] P. Haselager, R. Vlek, J. Hill, and F. Nijboer, "A note on ethical aspects of BCI," *Neural Networks*, 2009.

[12] N. K. N. Aznan, S. Bonner, J. D. Connolly, N. A. Moubayed, and T. P. Breckon, "On the Classification of SSVEP-Based Dry-EEG Signals via Convolutional Neural Networks," *arXiv:1805.04157*, may 2018. [Online]. Available: https://arxiv.org/abs/1805.04157 http://arxiv.org/abs/1805.04157

[13] Y. Wang, X. Chen, X. Gao, and S. Gao, "A Benchmark Dataset for SSVEP-Based Brain-Computer Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1746–1752, 2017.

[14] H. Cecotti, "A time-frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses," *Pattern Recognition Letters*, vol. 32, no. 8, pp. 1145–1153, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2011.02.022

[15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," pp. 1–21, 2016. [Online]. Available: http://arxiv.org/abs/1611.08024

[16] N. S. Kwak, K. R. Müller, and S. W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, pp. 1–20, 2017.

[17] I. Walker, M. Deisenroth, and A. Faisal, "Deep Convolutional Neural Networks for Brain Computer Interface using Motor Imagery," *Imperial College of Science, Technology and Medicine Department of Computing*, 2015. [Online]. Available: http://www.doc.ic.ac.uk/ mpd37/theses/DeepEEG_IanWalker2015.pdf

[18] Y. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, 2017.

[19] H. Cecotti and A. Gräser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[20] M. Liu, W. Wu, Z. Gu, Z. Yu, F. F. Qi, and Y. Li, "Deep learning based on Batch Normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, 2018.

[21] H. Cecotti, "Convolutional neural networks for event-related potential detection: Impact of the architecture," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2031–2034, 2017.

[22] M. Nakanishi, Y. Wang, Y. T. Wang, Y. Mitsukura, and T. P. Jung, "Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate," *PLoS ONE*, vol. 9, no. 6, 2014.

[23] 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. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/el.2010.0923

[24] M. Nakanishi, Y. Wang, Y. T. Wang, Y. Mitsukura, and T. P. Jung, "An approximation approach for rendering visual flickers in SSVEP-based BCI using monitor refresh rate," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013, pp. 2176–2179.

[25] F. Di Russo, S. Pitzalis, T. Aprile, G. Spitoni, F. Patria, A. Stella, D. Spinelli, and S. A. Hillyard, "Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential," *Human Brain Mapping*, 2007.

[26] Y. Wang, M. Nakanishi, Y. T. Wang, and T. P. Jung, "Enhancing detection of steady state visual evoked potentials using individual training data," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, 2014.

[27] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional Neural Networks for MatLab," in *Proceedings of the 23rd ACM international conference on Multimedia - MM '15*, 2015, pp. 689–692. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2733373.2807412

[28] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.