

PAPER • OPEN ACCESS

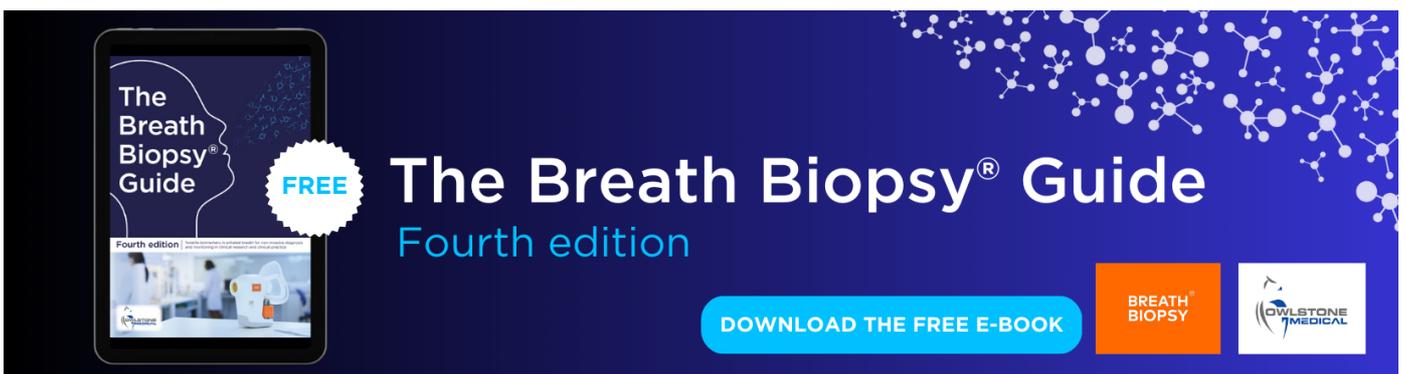
## Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials

To cite this article: Nicholas Waytowich *et al* 2018 *J. Neural Eng.* **15** 066031

View the [article online](#) for updates and enhancements.

You may also like

- [To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs](#)  
R Zerafa, T Camilleri, O Falzon et al.
- [The effect of distractors on SSVEP-based brain-computer interfaces](#)  
R Zerafa, T Camilleri, K P Camilleri et al.
- [The effect of stimulus number on the recognition accuracy and information transfer rate of SSVEP-BCI in augmented reality](#)  
Rui Zhang, Zongxin Xu, Lipeng Zhang et al.



The Breath Biopsy® Guide  
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

# Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials

Nicholas Waytowich<sup>1,2,5</sup> , Vernon J Lawhern<sup>1</sup> , Javier O Garcia<sup>1,3</sup>,  
Jennifer Cummings<sup>2</sup>, Josef Fallert<sup>2</sup>, Paul Sajda<sup>2</sup>  and Jean M Vettel<sup>1,3,4</sup>

<sup>1</sup> U S Army Research Laboratory, Aberdeen Proving Ground, MD, United States of America

<sup>2</sup> Laboratory for Intelligent Imaging and Neural Computing, Columbia University, New York, NY, United States of America

<sup>3</sup> Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States of America

<sup>4</sup> Department of Physiological Brain Sciences, University of California, Santa Barbara, CA, United States of America

E-mail: [nick.waytowich@gmail.com](mailto:nick.waytowich@gmail.com)

Received 12 March 2018, revised 22 August 2018

Accepted for publication 3 October 2018

Published 31 October 2018



CrossMark

## Abstract

*Objective.* Steady-state visual evoked potentials (SSVEPs) are neural oscillations from the parietal and occipital regions of the brain that are evoked from flickering visual stimuli. SSVEPs are robust signals measurable in the electroencephalogram (EEG) and are commonly used in brain–computer interfaces (BCIs). However, methods for high-accuracy decoding of SSVEPs usually require hand-crafted approaches that leverage domain-specific knowledge of the stimulus signals, such as specific temporal frequencies in the visual stimuli and their relative spatial arrangement. When this knowledge is unavailable, such as when SSVEP signals are acquired asynchronously, such approaches tend to fail. *Approach.* In this paper, we show how a compact convolutional neural network (Compact-CNN), which only requires raw EEG signals for automatic feature extraction, can be used to decode signals from a 12-class SSVEP dataset without the need for user-specific calibration. *Main results.* The Compact-CNN demonstrates across subject mean accuracy of approximately 80%, outperforming current state-of-the-art, hand-crafted approaches using canonical correlation analysis (CCA) and Combined-CCA. Furthermore, the Compact-CNN approach can reveal the underlying feature representation, revealing that the deep learner extracts additional phase- and amplitude-related features associated with the structure of the dataset. *Significance.* We discuss how our Compact-CNN shows promise for BCI applications that allow users to freely gaze/attend to any stimulus at any time (e.g. asynchronous BCI) as well as provides a method for analyzing SSVEP signals in a way that might augment our understanding about the basic processing in the visual cortex.

<sup>5</sup> Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: brain–computer interface, convolutional neural networks, steady-state visual evoked potentials, deep learning

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Evoked potentials are robust signals in the electroencephalogram (EEG) induced by sensory stimuli, and they have been used to study normal and abnormal function of the sensory cortex [1]. The most well-studied of these are steady-state visual evoked potentials (SSVEPs), which are neural oscillations in the visual cortex that are evoked from stimuli that temporally flicker in a narrow frequency band [2, 3]. SSVEPs likely arise from a reorganization of spontaneous intrinsic brain oscillations in response to a stimulus [4]. Paradigms leveraging SSVEP responses have been used to investigate the organization of the visual system [5, 6], identify biomarkers of disease and sensory function [7–9], and probe visual perception [10, 11].

The robustness of SSVEP has enabled its use as a control signal for brain computer interfaces (BCIs) that enable low-bandwidth communication for individuals with catastrophic loss of motor functions, bypassing neuro-muscular pathways and establishing a communication link directly to the brain [12, 13]. In a typical SSVEP BCI, a patient/subject is presented with a grid of squares on a computer monitor, where each square contains semantic information such as a letter, number, character, or action. Superimposed on these squares are visual flicker frequencies that uniquely ‘tag’ each square, thus mapping semantics to visual temporal frequency. As one directs their gaze and attention to a particular square with the semantic information they wish to convey, an SSVEP signal at the corresponding frequency can be measured in the EEG with dominant signals in parietal and occipital electrodes. The approach, though seemingly simple, has been important for enabling communication channels for those that are locked-in and have no other means of communication, particularly those with late stage amyotrophic-lateral sclerosis (ALS) [14].

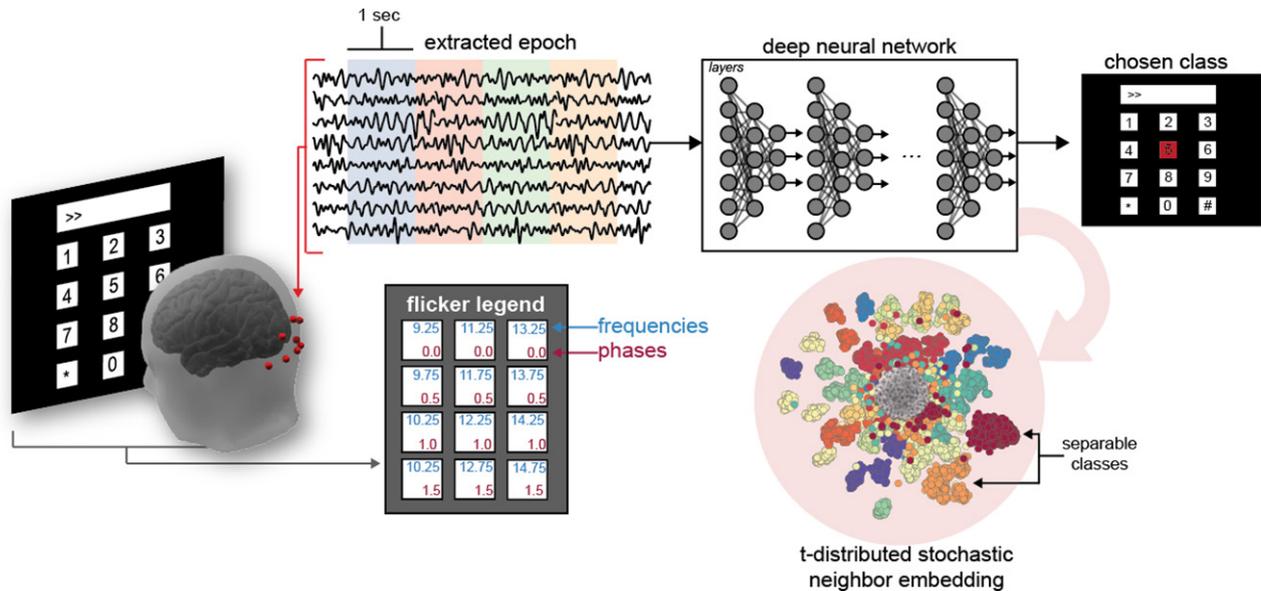
Central to using SSVEP, whether as a mechanism for probing vision or enabling a BCI, is the need to accurately decode and analyze the frequency information. Power spectral density analysis (PSDA), for example, is often used to identify spectral peaks in the EEG data that map to the flicker frequency of the stimulus. More recent approaches have used multivariate statistical analysis techniques, such as canonical correlation analysis (CCA), that employ a template matching scheme between the EEG data and a set of hand-crafted, sinusoidal reference signals [15–17]. A recent innovation in this approach, called Combined-CCA, uses a unique combination of sinusoidal templates as well as individual template responses constructed from SSVEP calibration data to improve the standard reference signals used in the CCA [18–20]. While Combined-CCA outperforms CCA, it does so at the cost of requiring user-specific calibration data to construct the reference signals.

To enable more flexible results without the need for *a priori* information, deep learning approaches were leveraged for their ability to learn robust feature representations. As such, deep learning techniques have surpassed traditional approaches that rely on manual feature extraction [21, 22]. Convolutional neural networks (CNNs) in particular have become a very popular deep learning approach for learning rich feature representations for image classification problems [23–26], and their ability to learn invariant features has shown promise to advance methods used in EEG signal analysis [27–35]. Deep learning approaches, however, typically require large amounts of training data in order to prevent over-fitting, and this requirement strongly limits their viability for SSVEP BCIs which are constrained by relatively modest sample sizes in typical BCI datasets. As a consequence, previous attempts at applying CNNs to SSVEP classification have used domain-specific representations to reduce the amount of training data required [36–38]. These approaches utilize the fast-fourier transform (FFT) in their deep-learning models, thereby transforming EEG signals from the time-domain to the frequency-domain. These FFT-based approaches constrain the model to learn only frequency-based features which may be insufficient to capture other task-relevant information. Consequently, this approach severely hinders the value of deep learning approaches on SSVEP datasets.

In this paper, we employ a deep learning approach that allows for the discovery of the underlying representations in SSVEP signals that relate flicker to semantics and train on relatively small datasets. Specifically, our approach utilizes a recently developed deep learning model by our group that is a compact convolutional neural network (Compact-CNN) and operates on broadly-filtered EEG signals. Its compact nature allows it to operate on smaller datasets, while the convolutional structure allows for the automatic extraction of task-relevant EEG features. Our Compact-CNN was applied to a previously collected SSVEP dataset [39] composed of 4 s long EEG epochs of data. Without using any user-specific calibration, our Compact-CNN results in substantially better classification accuracy compared to CCA and Combined-CCA. Furthermore, the underlying feature representations constructed by our Compact-CNN revealed that the deep learner is able to extract additional phase and amplitude related features associated with the SSVEP signals. We discuss how our Compact-CNN shows promise for BCI applications that allow users to be able to freely gaze/attend to any stimulus at any time (e.g. asynchronous BCI) and augments our understanding about the underlying cortical processing in the visual cortex.

## 2. Methods

Ten healthy participants volunteered for an offline SSVEP BCI experiment, and their de-identified data were downloaded



**Figure 1. Methodological overview.** Participants viewed a virtual keypad where each number flickered at a fixed frequency and phase (flicker legend). For each 4 s trial, the participant was cued to fixate on a specific number while EEG was recorded from occipital-parietal areas (left). The trials were then divided into 1 s segments and used as input to a deep learner, where subject-independent network models were trained and tested using a leave-one-subject-out classification procedure to identify the chosen class (the fixated number). Finally, the deep learning model activations were plotted using the t-SNE method to project the high-dimensional feature representation of the network down to two dimensions, revealing clusters that captured the diagnostic features within the test set (right).

from a publicly available repository [39]. The participants sat 60 cm away from a 27-inch LCD monitor (60 Hz refresh-rate and  $1280 \times 800$  resolution) in a dim room, and they looked at 12 flashing stimuli arranged in a  $4 \times 3$  grid of  $6 \text{ cm} \times 6 \text{ cm}$  squares that represented a numeric keypad. As shown in figure 1, the 12 SSVEP stimuli flashed at frequencies ranging from 9.25 Hz to 14.75 Hz in steps of 0.5 Hz. For each individual trial, a red square was used to cue subjects to visually fixate one of the 12 stimulus squares. Each trial was 4 s long and was within a block of 12 trials that included each of 12 unique targets. Each subject underwent 15 blocks of these 12 trials for a total of 180 trials. During the experiment, EEG data was collected from 8 active electrodes placed over occipital-parietal areas using the BioSemi ActiveTwo EEG system (Biosemi B.V., Netherlands) with a sampling frequency of 2048 Hz. All data were first bandpass filtered bidirectionally between 9 and 30 Hz using a Butterworth filter and then downsampled to 256 Hz. The 4 s epochs were divided into 1 s segments for subsequent analysis.

Three classification methods were compared using a leave-one-subject-out cross validation procedure where data from nine participants were pooled together for training and the 10th participant was used for testing. In this way, user-independent models were built that did not use any training data from the test participant. The test participant was rotated for each fold so that each participant became a test participant. For classification, three methods were compared that do not require user-specific calibration: our Compact-CNN and two baseline methods derived from the conventional multivariate statistical analysis technique, Canonical Correlation Analysis (CCA). The first method tested was the calibration-free CCA that used sinusoidal reference signals [15, 16], while the second was our variant of the state-of-the-art Combined-CCA

method [18, 19] that uses transfer learning to eliminate the user-specific calibration but maintains the superior classification performance of Combined-CCA [20].

Finally, the *learned* representation was visualized for our Compact-CNN, addressing a scientific aim within the EEG deep learning community to understand the diagnostic features of the data [32, 40]. As depicted in figure 1, the feature activations of the trained deep learner were shown using t-distributed stochastic neighbor embedding (t-SNE) to project individual SSVEP trials onto two dimensions [41]. The t-SNE projections were then plotted for each layer to visualize how the deep learner separates and clusters EEG trials in a projected feature-space learned by the deep network. The clusters are used to infer the diagnostic features within the test set.

## 2.1. Classification methods

### 2.1.1. EEGNet: compact convolutional neural network (Compact-CNN).

To assess the utility of deep learning approaches for SSVEP BCI, our analysis used our Compact-CNN (the EEGNet architecture) that was designed for classifying raw EEG when only limited amounts of data are available, such as in SSVEP BCI experiments [35]. Our Compact-CNN approach efficiently represents EEG signals in a compact manner by first performing temporal convolutions, with the convolutional kernel weights being identified from the data. The first layer of the network then performs a temporal convolution to mimic a bandpass frequency filter, a result supported by the convolution theorem [42]. Our approach then uses depthwise spatial convolutions that act as spatial filters to reduce the dimensionality of the data. The main benefit of depthwise convolutions is reducing the number of trainable parameters to fit, as these convolutions are not fully-connected to all

**Table 1.** Compact-CNN architecture, where  $C$  = number of channels,  $T$  = number of time points,  $F_1$  = number of temporal filters,  $F_2$  = number of separable filters (here,  $F_1 = F_2$ ),  $D$  = number of spatial filters to learn per temporal filter and  $N$  = number of classes, respectively.

| Layer           | Layer type      | # filters             | Size       | # params      | Output                           | Activation          | Options                                  |
|-----------------|-----------------|-----------------------|------------|---------------|----------------------------------|---------------------|--|
| 1               | Input           |                       |            |               | $(C, T)$                         |                     |  |
|                 | Reshape         |                       |            |               | $(1, C, T)$                      |                     |  |
|                 | Conv2D          | $F_1$                 | $(1, 256)$ | $256 * F_1$   | $(F_1, C, T)$                    | Linear              | Mode = same                              |
|                 | BatchNorm       |                       |            | $2 * F_1$     | $(F_1, C, T)$                    |                     |  |
|                 | DepthwiseConv2D | $D * F_1$             | $(C, 1)$   | $C * D * F_1$ | $(D * F_1, 1, T)$                | Linear              | Mode = valid, depth = D,<br>max norm = 1 |
|                 | BatchNorm       |                       |            | $2 * D * F_1$ | $(D * F_1, 1, T)$                |                     |  |
|                 | Activation      |                       |            |               | $(D * F_1, 1, T)$                | ELU                 |  |
|                 | AveragePool2D   |                       | $(1, 4)$   |               | $(D * F_1, 1, T // 4)$           |                     |  |
|                 | Dropout         |                       |            |               | $(D * F_1, 1, T // 4)$           |                     | Rate = 0.5                               |
|                 | 2               | SeparableConv2D       | $F_2$      | $(1, 16)$     | $16 * D * F_1 + F_2 * (D * F_1)$ | $(F_2, 1, T // 4)$  | Linear                                   |
| BatchNorm       |                 |                       |            | $2 * F_2$     | $(F_2, 1, T // 4)$               |                     |  |
| Activation      |                 |                       |            |               | $(F_2, 1, T // 4)$               | ELU                 |  |
| AveragePool2D   |                 |                       | $(1, 8)$   |               | $(F_2, 1, T // 32)$              |                     |  |
| Dropout         |                 |                       |            |               | $(F_2, 1, T // 32)$              |                     | Rate = 0.5                               |
| 3<br>Classifier |                 | Flatten               |            |               |                                  | $(F_2 * (T // 32))$ |  |
|                 | Dense           | $N * (F_2 * T // 32)$ |            |               | $N$                              | Softmax             |  |

previous outputs. When used in EEG-specific applications, this operation provides a direct way to learn spatial filters for each temporal filter, thus enabling the efficient extraction of frequency-specific spatial filters. The Compact-CNN also uses separable convolutions to more efficiently combine information across filters [43]. The main benefits of separable convolutions are (1) reducing the number of parameters to fit and (2) explicitly decoupling the relationship within and across outputs by first learning a kernel summarizing each output individually, then optimally merging the outputs afterwards. As described in table 1, each convolution layer is followed by batch normalization, 2D average pooling, and dropout layers. In the first two layers, the exponential linear unit (ELU) non-linearity (as opposed to other non-linear activation functions such as sigmoids or rectified linear units) was employed since it resulted in superior performance for EEG classification [35]. The fourth and final layer is connected to a dense layer with a softmax activation function for classification. Our Compact-CNN is trained using a categorical cross-entropy loss function (shown in equation (1)):

$$L_i = - \sum_j t_{ij} \log(p_{ij}) \quad (1)$$

where  $p$  are the model predictions,  $t$  are the true labels,  $i$  denotes the sample number, and  $j$  denotes the class. The model was implemented in Tensorflow [44], using the Keras API [45]. The model was trained for 500 iterations using the Adam [46] optimizer, with a minibatch size of 64 trials. The dropout probability was set to 0.5 for all layers. For this application, the model learned 96 temporal-spatial filter pairs ( $F_1 = 96$ ,  $F_2 = 96$  and  $D = 1$ ). The complete model architecture of our Compact-CNN is shown in table 1; our implementation can be found at <https://github.com/vlawhern/arl-eeemodels>.

**2.1.2. Standard CCA.** Canonical correlation analysis (CCA) is a statistical analysis technique that finds underlying correlations between two multidimensional datasets. Given two multidimensional variables  $X$  and  $Y$ , CCA seeks to find weight vectors  $W_x$  and  $W_y$ , such that their corresponding linear projections  $x = X^T W_x$  and  $y = Y^T W_y$  have maximal mutual correlation. These projections are found by solving the following objective function:

$$\max_{W_x, W_y} \rho(x, y) = \max_{W_x, W_y} \frac{E[W_x^T X Y^T W_y]}{\sqrt{E[W_x^T X X^T W_x] E[W_y^T Y Y^T W_y]}}. \quad (2)$$

For SSVEP detection, CCA computes the canonical correlation between multichannel EEG data and a set of reference signals  $Y_k$  composed of sines and cosines matching the fundamental and harmonic frequencies of the target stimulus. Reference signals are made for each of the  $K$  frequency stimuli where each set contains  $N_h$  harmonics of the fundamental frequency  $f_k$ .

$$Y_k = \begin{pmatrix} \sin(2\pi f_k t) \\ \cos(2\pi f_k t) \\ \vdots \\ \sin(2\pi N_h f_k t) \\ \cos(2\pi N_h f_k t) \end{pmatrix}. \quad (3)$$

EEG data is canonically correlated with each set of reference signals, and SSVEP detection is made on the basis of selecting the frequency set which provides the maximum canonical coefficient,  $f_k = \max_k \rho(x, f_k)$ , where  $f_k = f_1, f_2, \dots, f_K$ . In this analysis,  $K = 12$  since data was used from a 12-class SSVEP BCI [39] with numbers on the virtual keypad flashing at frequencies ranging from 9.25 Hz to 14.75 Hz in steps of 0.5 Hz.

**2.1.3. State-of-the-art Combined-CCA.** Combined-CCA is a recent extension to the CCA method that combines the reference signals from traditional CCA with prototype responses derived from participant EEG data to better account for the imperfect match between actual brain EEG data and the approximate sine and cosine signals [18, 19]. In the Combined-CCA method, participants complete a calibration session before using the SSVEP BCI to collect training data for each stimulus frequency, and EEG data for each frequency is averaged across the training trials. In our recent extension of Combined-CCA [20], we introduced a pooled transfer approach that averages SSVEP trial data from other subjects to eliminate a calibration session for the current user.

To create the prototype responses,  $\bar{X}_k$ , data were taken from 9 of the ten subjects and averaged across trials and subjects for each stimulus frequency  $k$ . Using these prototype responses  $\bar{X}_k$  as well as the sine and cosine reference signals  $Y_k$  (equation (3)), the Combined-CCA classifies the data from the test subject,  $X$ , by fusing together all pairs of canonical correlations between  $\bar{X}_k$ ,  $Y_k$  and  $X$  using a weighted average resulting in a combined correlation coefficient for each class,  $p_k$ . The frequency that maximizes this weighted correlation value is selected as the SSVEP target:

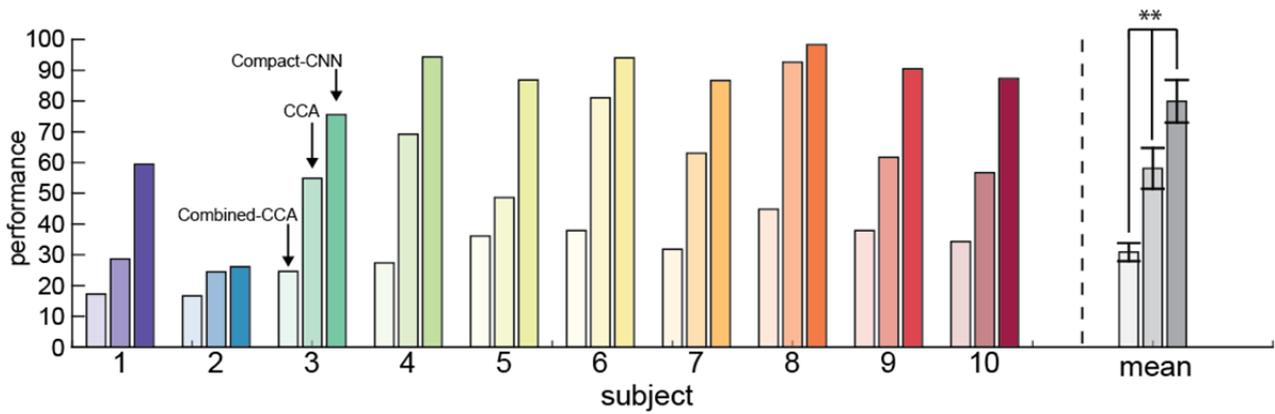
$$f_k = \max_k \rho(k), \quad k = 1, 2, \dots, K. \quad (4)$$

### 3. Results

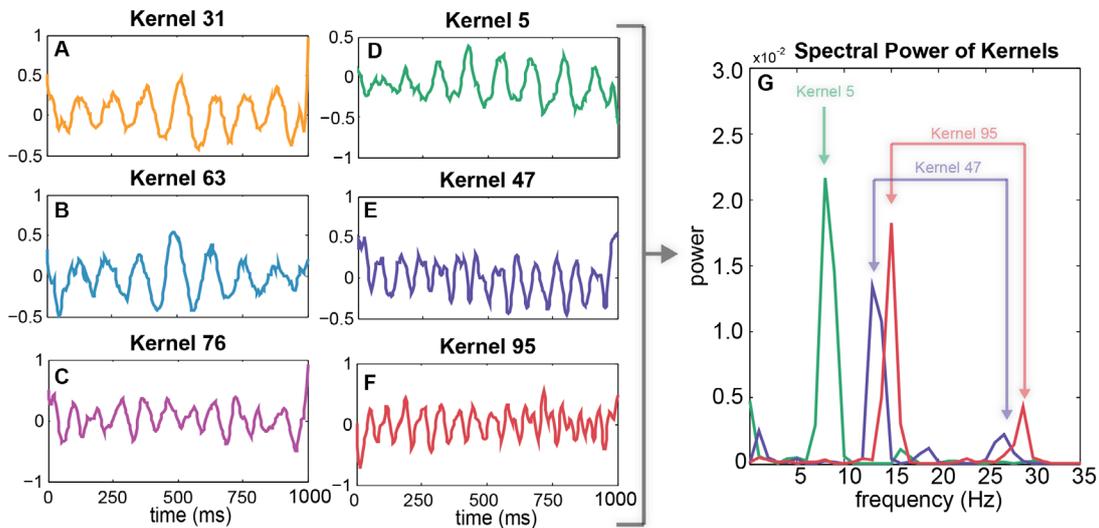
Our analysis compared three classification approaches on a 12-class SSVEP BCI experiment from a publicly available dataset [39]. Participants were cued to fixate one of 12 numbers flickering at frequencies ranging from 9.25 to 14.75 Hz on a virtual keypad for four seconds (figure 1). The 4 s epoch was divided into 1 s segments to augment the number of trials used in a 10-fold cross-validation procedure that identified the dominant frequency in the EEG and thus determined which number was fixated.

#### 3.1. Compact-CNN outperforms CCA and Combined-CCA

To assess whether deep learning approaches improve classification of SSVEP signals, performance of the Compact-CNN was compared to the multivariate statistical analysis technique, CCA, and its current state-of-the-art variant, Combined-CCA. The classification accuracy for each method is shown for each of the ten participants in figure 2, and results demonstrate the robust performance improvement for Compact-CNN relative to the two comparison methods. Compact-CNN proved to be particularly beneficial for subjects whose performance was



**Figure 2.** Classification accuracy. The SSVEP BCI classification accuracy is shown for each of ten participants for three methods, Combined-CCA, CCA, and Compact-CNN. The mean accuracy across participants is shown at the right, with significant paired t-test differences between Compact-CNN and each of the multivariate methods denoted with \*\*, representing a  $p < .0001$ . Error bars indicate SEM across participants. Nominal chance performance with 12 classes is 8.3%.



**Figure 3.** Visualization of Compact-CNN temporal kernels. (A–F) Representative subset of the derived temporal kernels. In the subplots, the x-axis denotes the length (in msec) of the temporal kernel, and the y-axis denotes the amplitude of the kernel. Kernels selected based on suspected frequency separability. (G) Spectral power of the temporal kernels shown in (D–F) that depict the separability of some kernels based on the frequency content of the temporal kernels. Figure labels denote the suspected stimulus driven responses (including harmonics, e.g., kernel 95, kernel 47).

notably poor when CCA based methods were used (i.e. subjects 1, 3, 9 and 10).

When evaluating the overall mean performance (figure 2, right), results demonstrated that Compact-CNN significantly outperformed CCA and Combined-CCA methods as indicated by paired t-tests (Compact-CNN versus Combined-CCA:  $t(9) = -10.5, p < 0.0001$ ; Compact-CNN versus CCA:  $t(9) = -8.7, p < 0.0001$ ). Surprisingly, the state-of-the-art Combined-CCA method performed significantly worse than CCA, and as subsequent analyses will show, a likely explanation is that the method is not prepared to work with asynchronous data that is not phase locked.

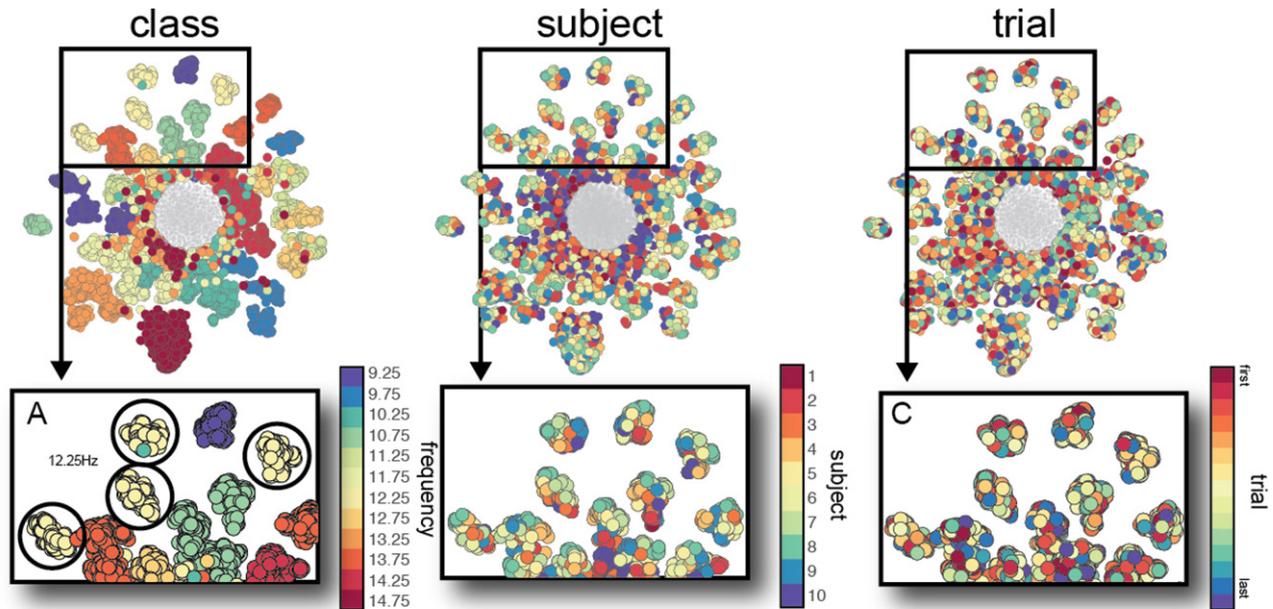
### 3.2. Compact-CNN extracts narrow-band frequency activity

Our analysis next examined the underlying learned representation of the Compact-CNN, investigating the diagnostic

features of the data that may account for the superior performance. Our first analysis along this effort focused on interpreting the learned temporal kernel weights in layer 1. Figure 3 visualizes a subset of the temporal kernels learned by the Compact-CNN model for one randomly-selected fold. Here we see that the Compact-CNN identifies narrow-band frequency activity along a spectrum of frequencies, both slow-wave (Kernel 63, at approx. 9 Hz) and fast-wave (Kernel 95, approx. 14 Hz). These frequencies closely align with that of the SSVEP experimental stimulus frequencies, suggesting that the model is capturing task-relevant oscillatory activity.

### 3.3. Compact-CNN reveals differences among classes

Our next analysis uses a data reduction and visualization technique (t-SNE) to investigate the hidden unit activation structure across all layers of the Compact-CNN. The activation in



**Figure 4.** *t*-SNE of Compact-CNN layer 3. Each element represents a single trial and the color represents membership for three different features: (A) the stimulus class of the 12-class SSVEP numeric keypad; (B) one of the ten participants; and (C) the trial number across the 180 trials. The bottom panels are zoomed-in sections of the top images to illustrate whether the individual clusters consist of the same colored label. Gray points indicate trials that are not visually separable and represent those least likely to be classified correctly.

each layer was similar, so only the *t*-SNE projections of the activation in layer 3 of the Compact-CNN were plotted in figure 4 for all training observations (across the nine training subjects) in fold 1.

This visualization is a projection of the data down to two dimensions, allowing us to estimate discriminability among training trials and to infer the properties of the learned representation that account for classification performance. Each element of the *t*-SNE plot represents a single trial with a color that indicates different known features. In figure 4, trials were labeled based on their class (one of 12 numbers on the virtual keypad), the participant ID (one of 10 unique individuals), and the trial order (one of 15 trial bins ordered from first to last). Only the class plot revealed coherent clusters of colored elements, indicating that the learner identified the 12 unique classes of the numeric keypad. The lack of clusters in the subject plot confirmed that the variability between individual participant brain signals does not account for the feature representation, and lack of clusters in the trial plot demonstrated that there was not a time-on-task effect (e.g. fatigue, drowsiness, inattention) that differentiated the trials across the duration of the experimental session.

Next, our analysis examined the composition of the clusters themselves. Interestingly, the class plot revealed separable clusters that were colored with the same class label. In figure 4(A), elements colored yellow belong to the trials where the participant fixated on the number 8 that flickered at 12.25 Hz. This reveals that the learner differentiated different trials from the same class. This unexpected separation is not predicted from the derivation of the model, as the objective function in equation (1) only seeks to separate the 12 classes. Consequently, our analysis investigated whether these within-class clusters may be related to another known feature of

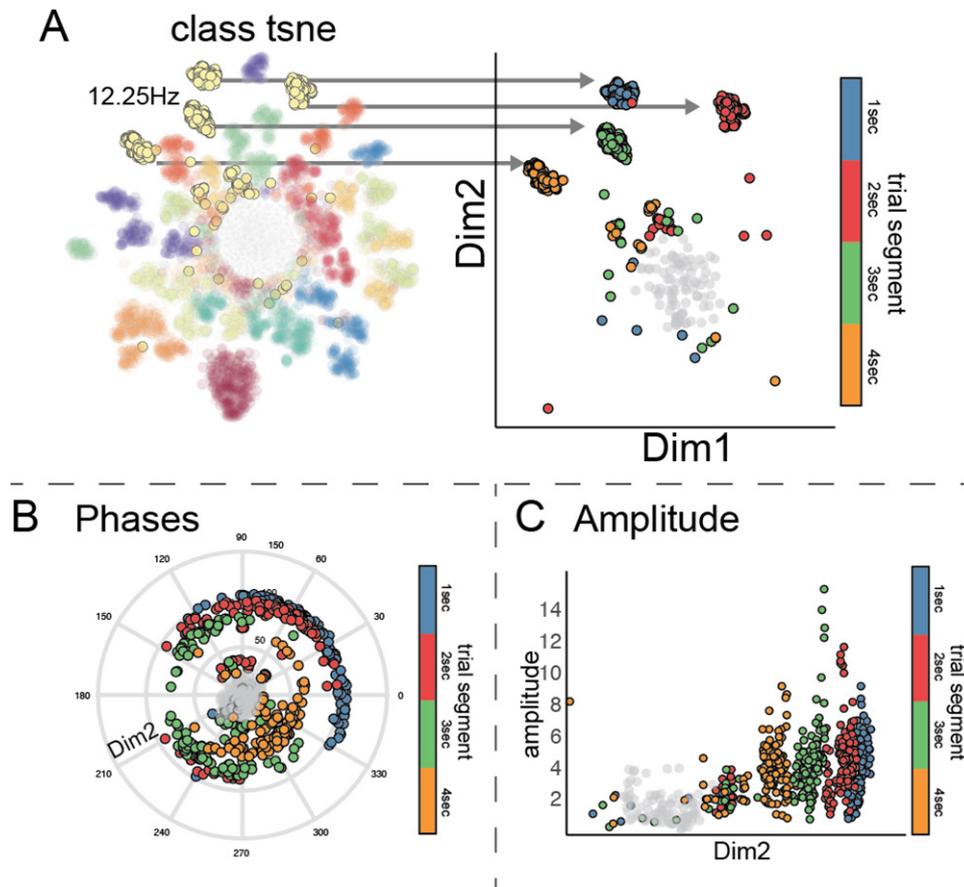
the dataset, namely the 1 s segments of the original 4 s trial epochs.

#### 3.4. Characterizing features of within-class clusters

In figure 5(A) (right), the elements in the 12.25 Hz class (yellow trial elements) were recolored according to their segment of the original 4 s trial, where 0–1 s is blue, 1–2 s is red, 2–3 s is green, and 3–4 s is orange. The segments account for the separable clusters within-class. This indicates that the deep learner is influenced (and learns) EEG features other than the trained class-level differences.

Next, our analysis characterized the signal properties of raw SSVEP responses for each trial with a phase and amplitude analysis shown in figures 5(B) and (C), respectively. The radial phase plot in figure 5(B) shows that the EEG trial segments have separable phases with the 3rd and 4th segments clustering at opposite phases to the 1st and 2nd segments of the trial. In contrast, figure 5(C) shows similar amplitudes across the four trial segments and only Dim2 values of figure 5(A) seem to denote trial segment separability.

While figure 5 illustrates phase and amplitude features for the 12.25 Hz class from channel Oz, the subplots of figure 6 confirm that these patterns are common across other stimulus classes and channels. The average estimated phase (left column) and amplitude (right column) are depicted for the 9.25 Hz class (top row), the 12.25 Hz class (middle row), and the 14.75 Hz class (bottom row). The bottom diagram in figure 6 shows the corresponding channel layout and coloring. Phase and amplitude estimations were averaged over each trial and subject for the corresponding stimulus frequency and are shown for each electrode as colored line plots. Segment-wise differences are strongest for phase, but they can also be



**Figure 5.** Phase and amplitude of 1 s trial segments. (A) The yellow 12.25 Hz clusters on the left are recolored on the right based on their respective 1 s segment of the original 4 s SSVEP trial, where blue is the first second, red the second, green third, and orange the fourth. (B) The polar angle plot of the estimated phase, where angle is represented in degrees within the circular plot and distance from the center of the plot represents the amplitude of Dim2 shown in (A). (C) The estimated amplitude of the phases of (B) plotted against the amplitude of Dim2 in (A). Data shown is from channel Oz which is located over the center of the visual cortex.

seen for amplitude features across all EEG channels. Here, only a few classes were illustrated for brevity, but this pattern is consistent across classes and scalp locations.

#### 4. Discussion

In this research, our analysis investigated whether deep learning approaches could improve performance in classification of SSVEP BCI trials over state-of-the-art methods. While CNNs have shown promise for feature extraction and classification of EEG signals across a number of domains [27–34], there has been very little work in applying deep learning to classification of SSVEP signals [36–38]. We demonstrate on a representative data set that our Compact-CNN approach can outperform both CCA and Combined-CCA. Our approach improved classification performance on a group level, but also in individuals who had particularly poor performance when conventional methods were used. An inspection of the *learned* representation used for classification in layer 3 of the Compact-CNN revealed activation clusters that differentiated the 12 classes of numeric stimuli in the virtual keypad. In contrast, the clusters did not differentiate individual participants from one another or the trial order that may arise from time-on-task fatigue, inattention, or other nuisance

variables. The specificity of the representation in layer 3 indicates that the superior performance was not dependent on task-irrelevant features (participant, time-on-task), indicating that the Compact-CNN did not merely overfit the data in the training phase. Instead, the superior performance of the Compact-CNN likely arose from diagnostic trial features that the deep learner identified were robust in the data, namely within-class separability.

Unexpectedly, the activation in layer 3 revealed clusters of trials from within the same class, even though the 12-class classification task did not require any within-class distinction. After further investigation of the trials in the clusters, results revealed that the variability in phase across the 1 s segments of the original 4 s SSVEP trial accounted for the within-class cluster differentiation, and the phase differences across the four 1 s segments were robust across channels and almost all of the stimulus classes. Although unexpected, these within-class clusters highlight the strength of the deep learning approaches to learn diagnostic features directly from the data.

Our original motivation for the 1 s segments arose from the aim to provide more training trials for the Compact-CNN. However, the segments led to two unexpected, but likely related, results: (1) the identification of phase variability across the EEG response in the original 4 s SSVEP trial

epoch and (2) the substantial performance impairment in classification accuracy for the current state-of-the-art method, Combined-CCA. In general, CCA approaches leverage domain knowledge to extract relevant SSVEP features embedded in the EEG responses, and our results suggest a limitation that may arise from assuming a particular stimulus frequency and phase. Most multivariate approaches, including Combined-CCA, have thus far only proven effective under system-paced or synchronized SSVEP paradigms where precise stimulus onset information is known ahead of time. Our results suggest that this approach may constrain the success of these methods to synchronous SSVEP paradigms, while our Compact-CNN approach shows promise for asynchronous SSVEP classification where the users would be able to freely gaze/attend to any stimulus at any time. Here, we discuss how our Compact-CNN approach demonstrates the value of deep learning for SSVEP BCI, provides a potential innovation for asynchronous BCIs, and augments our understanding about the underlying cortical processing in the visual cortex.

#### 4.1. Deep learning allows characterization of EEG features

Deep learning approaches allow computational models to learn representations of data with multiple layers of abstraction. This in turn can allow for the discovery of rich features or structure within large datasets. However, since these layers of features are learned instead of designed by human engineers, interpreting the meaning of these features remains a critical challenge in the field of deep learning [47, 48]. In this paper, our analysis utilized the representational learning capabilities of multilayer CNNs and show that the model learned interpretable features from EEG data that are consistent with the SSVEP literature. The Compact-CNN identifies phase features that are invariant to both subject and trial (time-on-task) differences.

Our results extend previous studies that have applied deep learning to SSVEP BCI paradigms. These prior approaches leveraged domain specific knowledge when constructing their feature extraction layers [36–38]. The first approach applied a CNN for the classification of SSVEP signals by using the fast-fourier transform (FFT) to convert time-domain representations into frequency-domain representations [36]. The second, more recent approach transformed SSVEP data into the frequency domain as a pre-processing step before classification with a CNN [37]. A similar approach was taken in [38], where the authors used an FFT-based hidden layer for classifying SSVEP signals. These approaches incorporate domain-specific knowledge of the stimulus frequencies by utilizing the FFT to transform raw time-domain EEG signals to the frequency domain since spectrally focal power differences are known to be discriminative features in the context of SSVEP-based decoding. All these approaches incorporate frequency specific knowledge, but no phase information. In contrast, our Compact-CNN operates on broadly-filtered EEG signals, and the relevant features (frequency and phase) are learned and extracted directly from these input signals. This

end-to-end approach might be particularly advantageous for use-cases where domain-specific knowledge is unavailable.

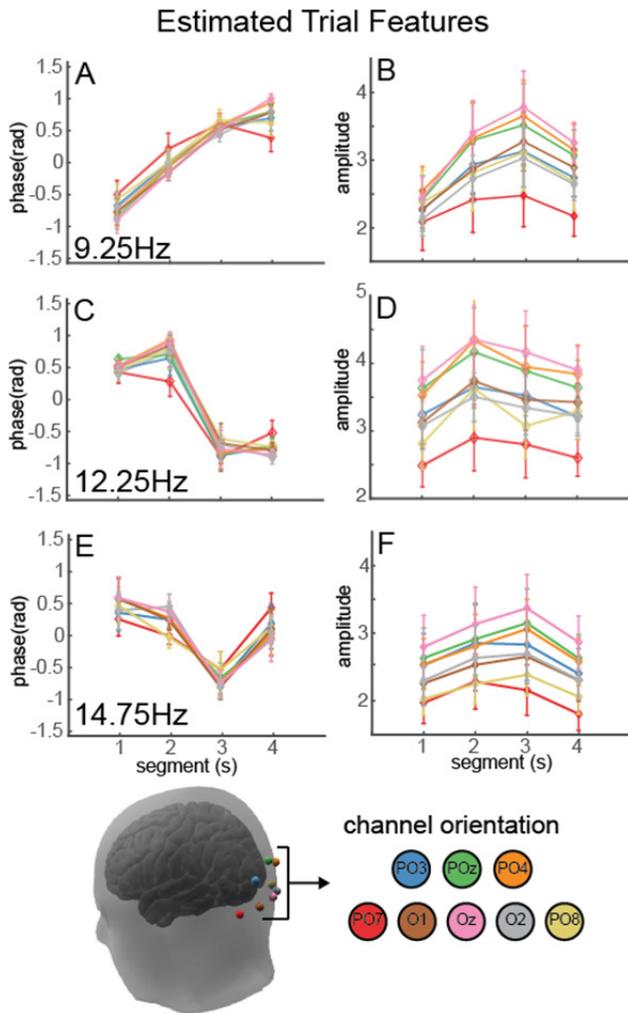
Additionally, our results demonstrated that the Compact-CNN was able to discover deeper structure within the SSVEP dataset in the form of phase information that reflects features of the 1 s segments of the original 4 s SSVEP trial epochs. This learned structure is present in the t-SNE projections of the model activations in the hidden layers. Thus, the Compact-CNN provides a powerful tool for interpreting intrinsic structures within EEG datasets, and critically, this type of structure discovery would not be possible using standard linear methods, such as CCA or Combined-CCA.

#### 4.2. Potential benefits of Compact-CNN-based asynchronous classification

SSVEP signals can be generated and classified using either synchronous or asynchronous temporal coding paradigms [49]. In synchronous paradigms, the onset and offset of the user gaze towards the stimulus are known: the user is cued to look at a flashing stimulus in a specified time-window so that the user's gaze is synchronized with the onset of stimulus flashing. In contrast, asynchronous SSVEP paradigms allow users to freely gaze/attend to any stimulus at any time. Asynchronous paradigms are generally more flexible and are more amenable to practical BCI. For some applications, asynchronous temporal coding paradigms can make BCI interaction faster, more intuitive, and/or ergonomic as the user can directly communicate a specific command of choice instead of waiting for the respective system cues in synchronized approaches. EEG-based decoding however, is typically more difficult in asynchronous SSVEP paradigms, particularly if phase information of the stimulus is not known.

Here, our division of the 4 s SSVEP trial epoch into 1 s segments simulates an asynchronous SSVEP BCI from a discretized stimulation paradigm. These segments approximate a less-controlled paradigm wherein a participant may randomly attend to a stimulus frequency. A segment length of 1 s represents a fast-paced BCI design where classifications can be made every second. On this data set, the Compact-CNN method outperformed state-of-the-art approaches in simulated asynchronous operation. The Compact-CNN learner extracted relevant frequency and phase features directly from the broadly-filtered EEG signals even when specific phase information of the SSVEP trials was unknown. This improvement could potentially benefit a variety of asynchronous SSVEP BCI applications including control of wheelchairs [17, 50], neuroprosthetics [51], exoskeletons [52], or interaction with virtual [53, 54] or augmented reality [55]. The observed performance improvement may also make this asynchronous approach attractive for applications like spellers [19], where synchronous paradigms have been traditionally preferred for their higher information bandwidth at the cost of lower flexibility.

Additionally, achieving calibrationless BCI classification is paramount for developing more practical BCI systems as it



**Figure 6.** Mean phase and amplitude across channels. Phase (left column) and amplitude (right column) are shown for three classes: the 9.25 Hz class ((A) and (B)), the 12.25 Hz class ((C) and (D)), and the 14.75 Hz class ((E) and (F)). Electrodes are represented as a separate line and mapped onto the scalp in the legend (bottom). Error bars indicate SEM across participants.

eliminates the need for time-consuming training sessions for the user [56]. Using a leave-one-subject out transfer protocol, our Compact-CNN approach outperformed the baseline CCA approaches without any user-specific calibration. This is further reflected in t-SNE projections which indicate that our deep learning model is able to learn subject invariant features, which is critical for transfer across subjects. Based on its promise for asynchronous BCI paradigms and the fact that user-specific calibration is not required, this Compact-CNN approach gives BCI designers additional options to tailor SSVEP BCI systems to the needs of specific user-groups or other requirements.

### 4.3. Detecting frequency and phase information with the Compact-CNN

The Convolution Theorem states that convolutions of signals in the time-domain relate to multiplication in the frequency-domain. Since the first layer of our Compact-CNN is a temporal convolution (whose weights need to be learned from

the data), our network has the capability to learn frequency-specific temporal filters, including EEG features as shown in our previous work [35]. We showed in figure 3 that the Compact-CNN model is capable of extracting narrow-band task-specific slow-wave and fast-wave frequency activity. We believe that our network is also capturing information correlated to frequency through the use of the average-pooling layer in layer 1 of our model. The sequence of operations in layer 1 (temporal convolution, ELU non-linearity then average-pooling) is similar to the methodology of [57] for calculating event-related synchronization and desynchronization features. In their work, they narrow-band filter, then square, then average over a moving window the signal to obtain an estimate of frequency power. The similarity of this approach to the operations in layer 1 of the Compact-CNN suggests that our model is calculating features at least correlated to that of frequency power.

In addition, convolutional neural networks have the property of equivariance to local translation [58–60]. In particular, when the data is a time-series signal, local translations correspond to phase offsets in the signal. This means that the output of the convolution operation (a dot product between the convolutional kernel and the signal) will preserve the phase information in the signal relative to the phase of the convolutional kernel (i.e. when the convolutional kernel is in phase or in anti-phase with the data). The ability of convolutional neural networks to capture translation-invariant features (in the case of a time-series signal, phase-invariant) was proven in [59]; it was also shown that pooling layers and model depth were critical to achieve this (additional theoretical treatment of convolutional neural networks and their ability to extract translation-invariant features can be found in [60]). It is through this property that our Compact-CNN approach was able to extract phase information from the signals of interest. Figure 5, which shows a t-SNE projection of the hidden unit activations in layer 3, provides additional evidence that our network can capture this phenomena.

### 4.4. Phase discrimination for vision neuroscience

While this paper largely focuses on the relevance for BCI applications, the Compact-CNN’s ability to detect phase holds incredible potential to augment our understanding of visual processing in common SSVEP experimental paradigms. For example, SSVEP experimentation has revealed the temporal dynamics and spatial constraints of spatial attention when divided across locations, refining theories of how we attend to specific locations in our environment [61]. Likewise, phase information about oscillations of alpha activity within coordinated populations of neurons in the visual cortex has been shown to predict whether stimuli will be detected in the environment, suggesting that phase information drives coordinated excitation that facilitates perception or synchronized inhibition that prevents stimulus detection [62–64].

Here, the Compact-CNN identified phase information in the SSVEP response that was irrelevant for the 12-class classification but robust in the neural response, both across stimulus frequencies as well as channels. The ability to detect nuanced

phase variability in EEG signals holds incredible potential to augment our understanding of the role of phase information in neural processing more generally. Variability in phase measured with EEG has been shown to modulate the neural response to a stimulus [65–68], influence reaction times [69, 70], and determine whether multisensory input is bound as a coherent percept [71]. Phase information within the neural response has been suggested to carry precise informational content [72, 73] and may even be a component of the neural code [74]. Our results demonstrate that a deep learner may provide an innovative way to reveal robust phase variability between stimuli that underlie the intricate coordination of neural activity that gives rise to cognition.

#### 4.5. Limitations

Within this dataset, the impaired performance of Combined-CCA likely arose from insufficient representation of the precise phase information across the 1s segments. Future research could investigate whether Combined-CCA could be adapted or used differently to better handle asynchronous operation. For example, provided that the BCI has knowledge of the phase of the stimulus, the template could be shifted so that its phase matches the phase of the stimulus. This may improve performance of Combined-CCA in an asynchronous setup. It is noteworthy though, that phase information of the stimulus is not available in all applications. Our Compact-CNN does not require phase information, suggesting its versatility for both experimental use and novel BCI applications.

Finally, our experimental setup did not include a non-control state, which is typically utilized in self-paced BCI systems. Future work could investigate the efficacy of our method in the presence of a non-control state in a closed-loop setting, which could lead to even more flexible and intuitive SSVEP-based human machine interaction.

#### Acknowledgments

This research was supported by mission funding to the US Army Research Laboratory and through Cooperative Agreement Number W911NF-10-2-0022. Support was also provided by the NSF under grant IIS-1527747. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

#### ORCID iDs

Nicholas Waytowich  <https://orcid.org/0000-0002-3786-0675>  
 Vernon J Lawhern  <https://orcid.org/0000-0002-3921-8723>  
 Paul Sajda  <https://orcid.org/0000-0002-9738-1342>

#### References

- [1] Regan D 1989 *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Amsterdam: Elsevier)
- [2] Herrmann C S 2001 Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena *Exp. Brain Res.* **137** 346–53
- [3] Regan D 1977 Steady-state evoked potentials *J. Opt. Soc. Am.* **67** 1475–89
- [4] Başar E 1980 *EEG-Brain Dynamics: Relation Between EEG and Brain Evoked Potentials* (Amsterdam: Elsevier)
- [5] Jeffreys A 1989 Book review on ‘*Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*’ by D Regan’ *Trends in Neurosciences* **12** 413–14
- [6] Silberstein R B, Nunez P L, Pipingas A, Harris P and Danieli F 2001 Steady state visually evoked potential (SSVEP) topography in a graded working memory task *Int. J. Psychophysiol.* **42** 219–32
- [7] Marx M, Bodis-Wollner I, Bobak P, Harnois C, Mylin L and Yahr M 1986 Temporal frequency-dependent VEP changes in parkinson’s disease *Vis. Res.* **26** 185–93
- [8] Clementz B A, Keil A and Kissler J 2004 Aberrant brain dynamics in schizophrenia: delayed buildup and prolonged decay of the visual steady-state response *Cogn. Brain Res.* **18** 121–9
- [9] Harding G, Wilkins A J, Erba G, Barkley G L and Fisher R S 2005 Photic-and pattern-induced seizures: expert consensus of the epilepsy foundation of america working group *Epilepsia* **46** 1423–5
- [10] Tononi G, Srinivasan R, Russell D P and Edelman G M 1998 Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses *Proc. Natl Acad. Sci.* **95** 3198–203
- [11] Garcia J O, Srinivasan R and Serences J T 2013 Near-real-time feature-selective modulations in human cortex *Curr. Biol.* **23** 515–22
- [12] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
- [13] Waytowich N R and Krusienski D J 2016 Multiclass steady-state visual evoked potential frequency evaluation using chirp-modulated stimuli *IEEE Trans. Hum. Mach. Syst.* **46** 593–600
- [14] Hsu H T, Lee I H, Tsai H T, Chang H C, Shyu K K, Hsu C C, Chang H H, Yeh T K, Chang C Y and Lee P L 2016 Evaluate the feasibility of using frontal SSVEP to implement an SSVEP-based BCI in young, elderly and ALS groups *IEEE Trans. Neural Syst. Rehabil. Eng.* **24** 603–15
- [15] Lin Z, Zhang C, Wu W and Gao X 2007 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs *IEEE Trans. Bio-Med. Eng.* **54** 1172–6
- [16] Bin G, Gao X, Yan Z, Hong B and Gao S 2009 An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method *J. Neural Eng.* **6** 046002
- [17] Waytowich N R and Krusienski D J 2017 Development of an extensible SSVEP-BCI software platform and application to wheelchair control *8th Int. IEEE/EMBS Conf. on Neural Engineering* pp 259–532
- [18] Nakanishi M, Wang Y-T Y, Mitsukura Y and Jung T-P 2014 A high-speed brain speller using steady-state visual evoked potentials *Int. J. Neural Syst.* **24** 1450019

- [19] Chen X, Wang Y, Nakanishi M, Gao X, Jung T-P and Gao S 2015 High-speed spelling with a noninvasive brain-computer interface *Proc. Natl Acad. Sci.* **12** 201508080
- [20] Waytowich N R, Faller J, Garcia J O, Vettel J M and Sajda P 2016 Unsupervised adaptive transfer learning for steady-state visual evoked potential brain-computer interfaces *Int. Conf. on Systems, Man, and Cybernetics*
- [21] Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Networks* **61** 85–117
- [22] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436
- [23] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *NIPS'12 Proc. 25th Int. Conf. on Neural Information Processing Systems* vol 1 1097–105 (<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>)
- [24] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *CoRR* (arXiv:1409.1556)
- [25] He K, Zhang X, Ren S and Sun J 2015 Deep residual learning for image recognition *CoRR* (arXiv:1512.03385)
- [26] Huang G, Liu Z, Weinberger K Q and van der Maaten L 2016 Densely connected convolutional networks *CoRR* (arXiv:1608.06993)
- [27] Antoniadis A, Spyrou L, Took C C and Sanei S 2016 Deep learning for epileptic intracranial EEG data *IEEE 26th Int. Workshop on Machine Learning for Signal Processing* pp 1–6
- [28] Liang J, Lu R, Zhang C and Wang F 2016 Predicting seizures from electroencephalography recordings: a knowledge transfer strategy *IEEE Int. Conf. on Healthcare Informatics* pp 184–91
- [29] Thodoroff P, Pineau J and Lim A 2016 Learning robust features using deep learning for automatic seizure detection *Machine Learning for Healthcare Conf.* pp 178–90
- [30] Mirowski P, Madhavan D, LeCun Y and Kuzniecky R 2009 Classification of patterns of {EEG} synchronization for seizure prediction *Clin. Neurophysiol.* **120** 1927–40
- [31] Långkvist M, Karlsson L and Loutfi A 2012 Sleep stage classification using unsupervised feature learning *Adv. Artif. Neural Syst.* **2012** 5
- [32] Schirrmester R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggenberger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Hum. Brain Mapp.* **38** 5391–420
- [33] Tabar Y R and Halici U 2017 A novel deep learning approach for classification of EEG motor imagery signals *J. Neural Eng.* **14** 016003
- [34] Gordon S M, Jaswa M, Solon A J and Lawhern V J 2017 Real world BCI: cross-domain learning and practical applications *Proc. of the 2017 ACM Workshop on an Application-oriented Approach to BCI out of the Laboratory (BCIforReal)* vol 17 (New York: ACM) pp 25–8
- [35] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 Eegnet: a compact convolutional neural network for EEG-based brain-computer interfaces *J. Neural Eng.* **15** 056013
- [36] Cecotti H 2011 A time-frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses *Pattern Recognit. Lett.* **32** 1145–53
- [37] Kwak N-S, Müller K-R and Lee S-W 2017 A convolutional neural network for steady state visual evoked potential classification under ambulatory environment *PLoS One* **12** 1–20
- [38] Bevilacqua V, Tattoli G, Buongiorno D, Loconsole C, Leonardi D, Barsotti M, Frisoli A and Bergamasco M 2014 A novel BCI-SSVEP based approach for control of walking in virtual environment using a convolutional neural network *Int. Joint Conf. on Neural Networks* pp 4121–8
- [39] Nakanishi M, Wang Y, Wang Y-T and Jung T-P 2015 A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials *PLoS One* **10** e0140703
- [40] Sturm I, Lapuschkin S, Samek W and Müller K-R 2016 Interpretable deep neural networks for single-trial EEG classification *J. Neurosci. Methods* **274** 141–5
- [41] Maaten L V D and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605
- [42] Cohen M X 2014 *Analyzing Neural Time Series Data: Theory and Practice* (Cambridge, MA: MIT Press)
- [43] Chollet F 2016 Xception: deep learning with depthwise separable convolutions *CoRR* (arXiv:1610.02357)
- [44] Abadi M et al 2016 Tensorflow: a system for large-scale machine learning *Proc. of the 12th USENIX Conf. on Operating Systems Design and Implementation (Berkeley: USENIX Association)* pp 265–83
- [45] Chollet F et al 2015 Keras <https://github.com/keras-team/keras>
- [46] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
- [47] Goodfellow I J, Shlens J and Szegedy C 2014 Explaining and harnessing adversarial examples (arXiv:1412.6572)
- [48] Nguyen A M, Yosinski J and Clune J 2014 Deep neural networks are easily fooled: high confidence predictions for unrecognizable images *CoRR* (arXiv:1412.1897)
- [49] Mason S G, Bashashati A, Fatourehchi M, Navarro K F and Birch G E 2007 A comprehensive survey of brain interface technology designs *Ann. Biomed. Eng.* **35** 137–69
- [50] Diez P F, Müller S M T, Mut V A, Laciari E, Avila E, Bastos-Filho T F and Sarcinelli-Filho M 2013 Commanding a robotic wheelchair with a high-frequency steady state visual evoked potential based brain-computer interface *Med. Eng. Phys.* **35** 1155–64
- [51] Müller-Putz G R, Scherer R, Pfurtscheller G and Rupp R 2006 Brain-computer interfaces for control of neuroprostheses: from synchronous to asynchronous mode of operation *Biomed. Tech.* **51** 57–63
- [52] Kwak N S, Müller K-R and Lee S W 2015 A lower limb exoskeleton control system based on steady state visual evoked potentials *J. Neural Eng.* **12** 056009
- [53] Legény J, Abad R V and Lécuyer A 2011 Navigating in virtual worlds using a self-paced SSVEP-based brain-computer interface with integrated stimulation and real-time feedback *Presence: Teleoperators Virtual Environ.* **20** 529–44
- [54] Waytowich N R, Yamani Y and Krusienski D J 2017 Optimization of checkerboard spatial frequencies for steady-state visual evoked potential brain-computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 557–65
- [55] Faller J, Allison B Z, Brunner C and Scherer R 2010 A feasibility study on SSVEP-based interaction with motivating and immersive virtual and augmented reality *Technical Report* Graz University of Technology
- [56] Waytowich N R, Lawhern V J, Bohannon A W, Ball K R and Lance B J 2016 Spectral transfer learning using information geometry for a user-independent brain-computer interface *Frontiers Neurosci.* **10** 430
- [57] Pfurtscheller G and da Silva F L 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57
- [58] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
- [59] Wiatowski T and Bölcskei H 2018 A mathematical theory of deep convolutional neural networks for feature extraction *IEEE Trans. Inf. Theory* **64** 1845–66

- [60] Mallat S 2016 Understanding deep convolutional networks *Phil. Trans. R. Soc. A* **374** (<http://rsta.royalsocietypublishing.org/content/374/2065/20150203>)
- [61] Itthipuripat S, Garcia J O and Serences J T 2013 Temporal dynamics of divided spatial attention *J. Neurophysiol.* **109** 2364–73
- [62] Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann C S and Bäuml K-H 2007 Prestimulus oscillations predict visual perception performance between and within subjects *NeuroImage* **37** 1465–73
- [63] Mathewson K E, Gratton G, Fabiani M, Beck D M and Ro T 2009 To see or not to see: prestimulus  $\alpha$  phase predicts visual awareness *J. Neurosci.* **29** 2725–32
- [64] Brooks J R, Garcia J O, Kerick S E and Vettel J M 2016 Differential functionality of right and left parietal activity in controlling a motor vehicle *Frontiers Syst. Neurosci.* **10** 106
- [65] Barry R J, Rushby J A, Johnstone S J, Clarke A R, Croft R J and Lawrence C A 2004 Event-related potentials in the auditory oddball as a function of EEG alpha phase at stimulus onset *Clin. Neurophysiol.* **115** 2593–601
- [66] Makeig S, Westerfield M, Jung T-P, Enghoff S, Townsend J, Courchesne E and Sejnowski T J 2002 Dynamic brain sources of visual evoked responses *Science* **295** 690–4
- [67] Jansen B H and Brandt M E 1991 The effect of the phase of prestimulus alpha activity on the averaged visual evoked response *Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect.* **80** 241–50
- [68] Haig A R and Gordon E 1998 Prestimulus EEG alpha phase synchronicity influences n100 amplitude and reaction time *Psychophysiology* **35** 591–5
- [69] Callaway E and Yeager C L 1960 Relationship between reaction time and electroencephalographic alpha phase *Science* **132** 1765–6
- [70] Dustman R E and Beck E C 1965 Phase of alpha brain waves, reaction time and visually evoked potentials *Electroencephalogr. Clin. Neurophysiol.* **18** 433–40
- [71] van Atteveldt N, Murray M M, Thut G and Schroeder C E 2014 Multisensory integration: flexible use of general operations *Neuron* **81** 1240–53
- [72] Hopfield J J 1995 Pattern recognition computation using action potential timing for stimulus representation *Nature* **376** 33
- [73] Jensen O and Lisman J E 2000 Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding *J. Neurophysiol.* **83** 2602–9
- [74] Perkel D H and Bullock T H 1968 Neural coding *Neurosci. Res. Program Bull.* **6** 221–348