SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Brain Machine Interfaces and Applications

April 2022, Vol. 65 140406:1–140406:13
https://doi.org/10.1007/s11432-022-3440-5

SSVEP-based brain-computer interfaces are vulnerable to square wave attacks

Rui BIAN[1], Lubin MENG[1] & Dongrui WU[1,2*]

[1]*School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan* 430074*, China;*
[2]*Zhejiang Lab, Hangzhou* 311121*, China*

Received 2 January 2022/Revised 14 February 2022/Accepted 28 February 2022/Published online 14 March 2022

Abstract Electroencephalogram (EEG) based brain-computer interfaces (BCIs) have attracted wide attention in recent years. Steady-state visual evoked potential (SSVEP) is one of the most popular BCI paradigms, which has high information transmission rate and short user calibration time. Recent research has shown that EEG-based BCIs are under the threat of adversarial examples. However, existing attack approaches are very difficult to implement in a real-world system. Considering SSVEP's dependency on the frequency information, this paper proposes to use square wave signals as adversarial perturbations to attack SSVEP-based BCIs, which are easy to generate and apply in practice. EEG trials contaminated by the square wave perturbation can be classified into any target class specified by the attacker. Compared with previous approaches, our perturbation can be implemented much more easily. Experiments on two SSVEP datasets demonstrated the efficiency and robustness of the proposed approach in attacking two SSVEP classifiers based on canonical correlation analysis, exposing a critical security problem in SSVEP-based BCIs.

Keywords electroencephalogram, brain-computer interface, steady-state visual evoked potential, adversarial attack

Citation Bian R, Meng L B, Wu D R. SSVEP-based brain-computer interfaces are vulnerable to square wave attacks. Sci China Inf Sci, 2022, 65(4): 140406, https://doi.org/10.1007/s11432-022-3440-5

1 Introduction

A brain-computer interface (BCI) offers the user a direct communication channel between the brain and an external device, such as a computer, wheelchair, robot, etc. [1]. Electroencephalogram (EEG) [2], which records the brain electrical activities from the scalp, has become the most popular input signal in BCIs, due to its simplicity and low cost.

There are several different paradigms in EEG-based BCIs, e.g., P300 evoked potential [3–7], motor imagery (MI) [8,9], steady-state visual evoked potential (SSVEP) [10], etc. Compared with other paradigms, SSVEP usually has a higher information transmission rate and shorter user calibration time, so it has been widely used in BCI spellers [11] and device controllers [12].

An SSVEP is the neural response to a visual stimulus at a specific frequency. When the user stares at a target flickering at a frequency ranging from 3.5 to 75 Hz, the brain generates EEG signals at the same (or multiples of the) frequency of the target [13]. Thus, a classifier only needs to identify the dominant frequency of the user's EEG signal to decode which target the user is paying attention to.

Multiple machine learning approaches [14, 15] have been developed to improve the performance of SSVEP-based BCIs, such as canonical correlation analysis (CCA) [16], filter bank CCA (FBCCA) [17], task-related component analysis (TRCA) [18]. Their goal is to make SSVEP-based BCIs faster and more accurate; however, few studies have considered the security of SSVEP-based BCIs.

Refs. [19, 20] have shown that machine learning models are vulnerable to adversarial examples. Adversarial examples are normal samples contaminated by deliberately designed tiny perturbations, which are hard to be noticed by human eyes but can fool machine learning models easily. Adversarial examples have been found in many application domains. Szegedy et al. [19] first discovered adversarial

* Corresponding author (email: drwu@hust.edu.cn)

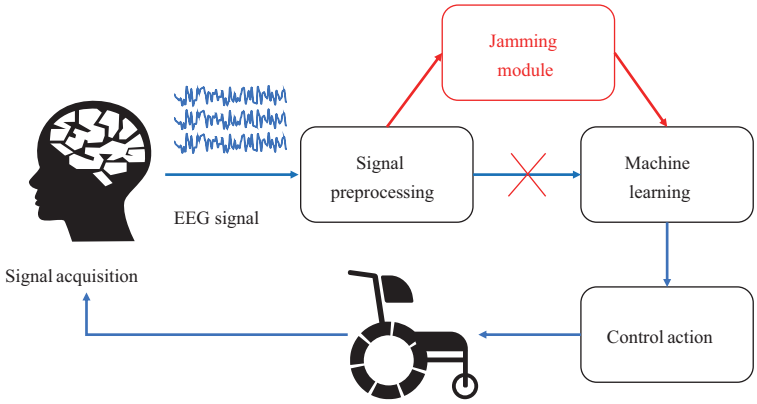


Figure 1 (Color online) Attack an EEG-based BCI system using a jamming module [23].

examples in image classification. Carlini and Wagner [21] demonstrated that adversarial examples exist in the speech by attacking DeepSpeech, a state-of-the-art speech-to-text transcription neural network. Han et al. [22] found that deep learning models for electrocardiogram classification are susceptible to adversarial attacks. Zhang and Wu [23] showed that machine learning models in EEG-based BCIs are also vulnerable to adversarial attacks. Many attack approaches have been proposed to achieve a higher attack success rate (ASR), e.g., fast gradient sign method (FGSM) [20], DeepFool [24], projected gradient descent (PGD) [25].

According to how adversarial attacks are performed, they can be grouped into two types. One is the evasion attack [20, 24, 25], in which the attacker attacks a machine learning model by adding tiny perturbations to the benign test samples, i.e., contaminating the test set. Another is poisoning attack [26, 27], where the attacker contaminates the training set to insert a backdoor to the trained model. By adding the backdoor key to a test sample, the attacker can manipulate the model's output easily.

It has been shown that both types of attacks exist in EEG-based BCIs. Zhang and Wu [23] were the first to point out that deep learning models in EEG-based BCIs are vulnerable to adversarial attacks. They successfully attacked three popular convolutional neural network (CNN) classifies (EEGNet [28], DeepCNN [29], and ShallowCNN [29]) using FGSM-based approaches. However, their approaches for computing the perturbation need to know the entire EEG trial, which is theoretically important but difficult to implement in real-time BCIs, because once the entire EEG trial is observed, it has already been sent out, and there is no way to add the perturbation to it. In other words, their approach does not satisfy causality.

To solve this problem, Liu et al. [30] introduced universal adversarial perturbation to EEG-based BCIs, which generates a fixed perturbation for all test samples. They proposed a novel total loss minimization approach to generate the universal adversarial perturbation. Recently, Zhang et al. [31] used gradient-based approaches to generate adversarial perturbation templates to attack traditional machine learning models in P300 and SSVEP spellers.

The above two approaches satisfy causality, but they still have some limitations. First, their adversarial perturbations are very complex and vary by EEG channels, which may not be easy to generate in practice. Second, both of them use the same framework as shown in Figure 1, where a jamming module is injected between signal preprocessing and machine learning to add perturbations to the test samples. For many BCIs, since both blocks are integrated together, it may not be easy to inject the jamming module.

Recently, Meng et al. [32] successfully performed a poisoning attack in EEG-based BCIs using a narrow period pulse as the backdoor key. Due to the simplicity of the narrow period pulse, their approach is more practical than all previous ones. It can easily contaminate the test samples while EEG signals are being collected because the backdoor key has been determined in advance.

This paper proposes an even easier and more practically realizable evasion attack approach for SSVEP-based BCIs. Because of SSVEP's dependency on the frequency information, a square wave signal is used as our perturbation, which can significantly disturb the frequency information of EEG trials. What is more, the square wave signal can be easily generated, making the proposed attack simple to implement.

Our main contributions are as follows.

(1) We showed, for the first time, that SSVEP-based BCIs can be attacked by perturbations with a specific frequency, which exposes a critical security problem in SSVEP-based BCIs.

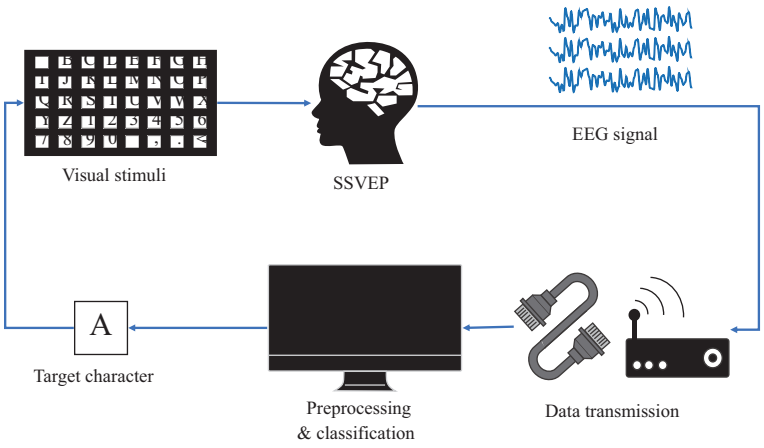


Figure 2 (Color online) An SSVEP speller.

(2) We proposed an easy and practically realizable evasion attack approach in SSVEP-based BCIs. We chose a square wave signal as the perturbation, which can be easily generated by a signal generator.

(3) We demonstrated the effectiveness of the proposed attack approach on two SSVEP datasets, and the ASRs on the CCA-based models were almost 100%. We also verified that the proposed attack approach can resist EEG preprocessing and is insensitive to the phase of the square wave signal.

The remainder of this paper is organized as follows. Section 2 introduces our proposed square wave attack approach. Section 3 presents the details of the experimental setting and results. Section 4 draws the conclusion and points out several future research directions.

2 Square wave attack

This section introduces first SSVEP spellers, then the victim classifiers in them, and finally our proposed square wave attack approach.

2.1 SSVEP spellers

A classical SSVEP speller is shown in Figure 2. There are 40 characters flickering at specific frequencies (usually ranging from 8 to 15.8 Hz with 0.2 Hz increment) on the screen. The user needs to stare at a target character, while his/her EEG signals are being collected by an EEG headset and transmitted through wires or wirelessly to a computer for analysis. When a long enough EEG trial is collected, an algorithm processes the EEG trial to mine its frequency information. Finally, the target character can be identified.

2.2 The victim models

This subsection introduces two CCA-based models, which are widely used in SSVEP-based BCIs due to their simplicity and training-free nature. They are the victim of our proposed square wave attack approach.

CCA. Lin et al. [16] proposed to use CCA to enhance the signal-to-noise ratio (SNR) of SSVEPs. CCA-based frequency recognition approaches have achieved high information transmission rates [33, 34]. CCA extracts frequency information of SSVEPs by calculating the canonical correlation coefficients between the EEG signals and the standard reference signals, which consist of sinusoidal signals of a stimulation frequency and its harmonics. Its main idea is to find linear combinations of the two signals to maximize their correlation.

Let $X \in \mathbb{R}^{C\times S}$ be an EEG trial, where C is the number of EEG channels and S is the number of time-domain samples, R_f be a standard reference signal for the stimulation frequency f, and f_s be the

sampling rate. Let

$$R_f = \begin{bmatrix} \sin\left(\frac{2\pi f}{f_s}\right) & \sin\left(\frac{4\pi f}{f_s}\right) & \cdots & \sin\left(\frac{2S\pi f}{f_s}\right) \\ \cos\left(\frac{2\pi f}{f_s}\right) & \cos\left(\frac{4\pi f}{f_s}\right) & \cdots & \cos\left(\frac{2S\pi f}{f_s}\right) \\ \sin\left(\frac{4\pi f}{f_s}\right) & \sin\left(\frac{8\pi f}{f_s}\right) & \cdots & \sin\left(\frac{4S\pi f}{f_s}\right) \\ \cos\left(\frac{4\pi f}{f_s}\right) & \cos\left(\frac{8\pi f}{f_s}\right) & \cdots & \cos\left(\frac{4S\pi f}{f_s}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \sin\left(\frac{2\pi nf}{f_s}\right) & \sin\left(\frac{4\pi nf}{f_s}\right) & \cdots & \sin\left(\frac{2S\pi nf}{f_s}\right) \\ \cos\left(\frac{2\pi nf}{f_s}\right) & \cos\left(\frac{4\pi nf}{f_s}\right) & \cdots & \cos\left(\frac{2S\pi nf}{f_s}\right) \end{bmatrix}, \tag{1}$$

where n is the number of harmonics ($n = 4$ in this paper).

The goal of CCA is to find weight vectors W_X and W_{R_f}, and the maximum canonical correlation coefficient $c(X, R_f)$, by solving the following problem:

$$\begin{aligned} c(X, R_f) &= \max_{W_X, W_{R_f}} \rho\left(X^{\mathrm{T}} W_X, R_f^{\mathrm{T}} W_{R_f}\right) \\ &= \max_{W_X, W_{R_f}} \frac{\mathrm{E}[W_X^{\mathrm{T}} X R_f^{\mathrm{T}} W_{R_f}]}{\sqrt{\mathrm{E}\left[W_X^{\mathrm{T}} X X^{\mathrm{T}} W_X\right] \mathrm{E}[W_{R_f}^{\mathrm{T}} R_f R_f^{\mathrm{T}} W_{R_f}]}}, \end{aligned} \tag{2}$$

where ρ is the correlation coefficient, superscript T is matrix or vector transpose, and E is the expectation.

Let K be the number of targets ($K = 40$ in this paper), and f_k be the flickering frequency of the k-th target. The stimulation frequency f^* of an SSVEP trial is determined by

$$f^* = \arg\max_{f_k} c(X, R_{f_k}). \tag{3}$$

FBCCA. Chen et al. [17] first integrated the filter bank technique and CCA to enhance the classification accuracy of SSVEP-based BCIs.

In FBCCA, the original EEG signal X is decomposed into M ($M = 5$ in this paper) sub frequency bands $\{X_m\}_{m=1}^{M}$ by M different bandpass filters. All filters have the same upper cut-off frequency of 88 Hz, but the m-th filter has a lower cut-off frequency of $8m$ Hz.

Then, a separate CCA is performed between each sub-band component and the reference signals. A weighted sum of the squares of those correlation coefficients is calculated as the feature for target classification. The weight for the m-th sub-band is [17]

$$w(m) = m^{-1.25} + 0.25, \quad m = 1, 2, \ldots, M. \tag{4}$$

The stimulation frequency f^* is determined by the largest sum of the weighted squares of the canonical correlation coefficients, i.e.,

$$f^* = \arg\max_{f_k} \sum_{m=1}^{M} w(m) c^2(X_m, R_{f_k}). \tag{5}$$

2.3 The attack approach

In previous approaches to attack a BCI system, the attacker needs to know its machine learning model and observe the entire test EEG trial to generate an adversarial perturbation. Even using universal adversarial perturbations, the attacker has to know the start time of the EEG trial to inject the perturbation. However, these may be difficult to implement. The data preprocessing and machine learning modules are usually integrated together, so it is very challenging to inject a jamming module between them. More importantly, for a real-world causal system, the attacker should add the perturbation while the test EEG trial is being transmitted, instead of after it has been transmitted.

Considering SSVEP's dependency on the frequency information, we propose an attack framework using square wave signals as the perturbations. As shown in Figure 3, an 8 Hz square wave signal is composed

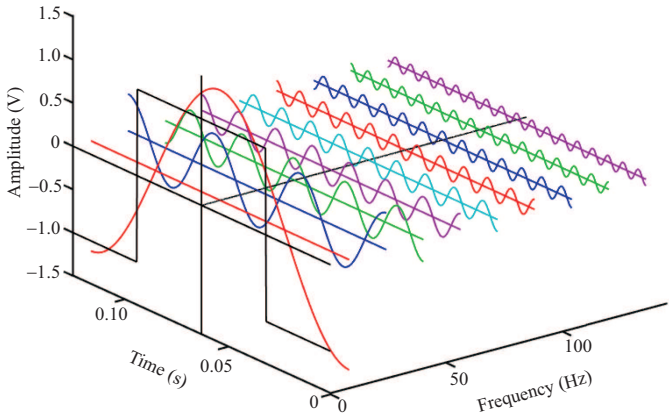


Figure 3 (Color online) Fourier transform of an 8 Hz square wave signal, showing only the odd harmonics.

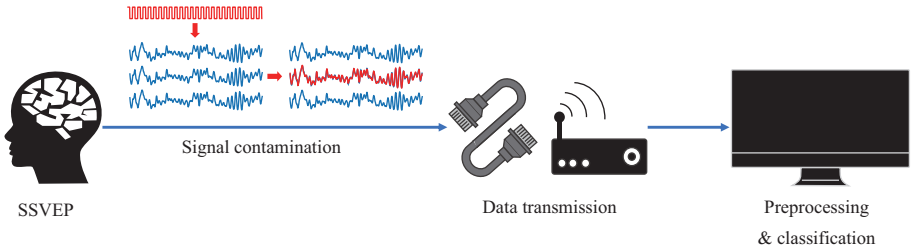


Figure 4 (Color online) Our proposed square wave attack framework.

of an 8 Hz sinusoidal signal and its harmonic signals (only the odd harmonics are shown). Adding such a square wave signal to the EEG trial enhances the energy of these frequencies. In addition, square wave signals are very simple and can be generated easily by a signal generator.

More specifically, to attack an SSVEP-based BCI system with K targets corresponding to K frequencies, the attacker has to prepare K square wave perturbations with the same frequencies. To change the SSVEP output to a target with a specific frequency, the attacker only needs to add a square wave perturbation with the same frequency to the test sample.

A square wave perturbation with frequency f, amplitude a, and phase ϕ is defined as

$$S(t) = a \cdot \operatorname{sign}(\sin(2\pi f t + \phi)). \tag{6}$$

The attack framework is shown in Figure 4. It directly adds square wave perturbations to EEG trials before the transmission. Experiments show that this attack approach is insensitive to the phase of the square wave perturbation, which means the attack can start at any time when the BCI system is used. These characteristics make our attack approach very practical in real-world BCI systems.

In summary, our proposed square wave attack approach has three main advantages.

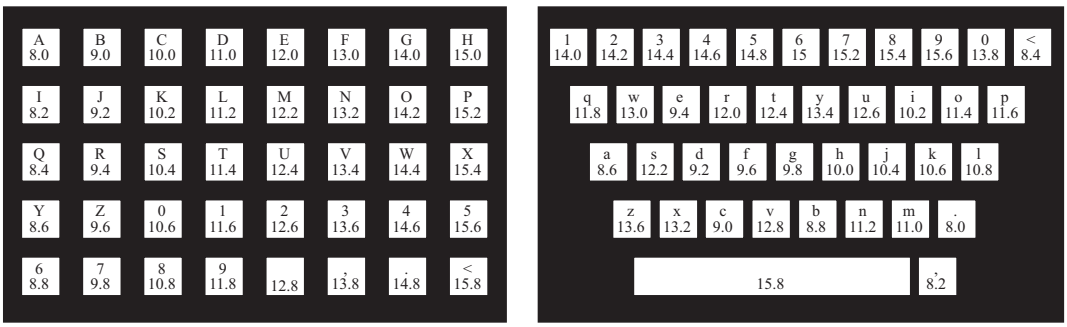
(1) The attacker only needs to know very little information about the SSVEP speller, i.e., the flickering frequency of each target, and about the user, i.e., the standard deviation (std.) of his/her EEG signals to quickly set an appropriate amplitude for the perturbation. However, the latter is not mandatory: the attacker can always start from a small amplitude and gradually increase it to achieve the best trade-off between ASRs and stealthiness.

(2) Unlike the complex waveforms generated by other approaches [30, 31], square wave perturbations are very simple to generate.

(3) Square wave perturbations can be easily added to EEG trials in real-time, for example, by attaching a signal generator to one or more specific channels.

3 Experiments

This section introduces the experimental settings and results to validate the effectiveness of our proposed square wave attack approach.



(a) (b)

Figure 5 Stimulation interface of (a) Benchmark and (b) BETA.

3.1 SSVEP datasets

The following two publicly available SSVEP datasets were used in our experiments.

Benchmark. The Benchmark SSVEP dataset, acquired using a 40-target BCI speller, was first introduced by Wang et al. [35] in 2017. It consists of 64-channel EEG signals collected from 35 healthy subjects. As shown in Figure 5(a), 40 targets flicker at different frequencies (ranging from 8 to 15.8 Hz with an interval of 0.2 Hz) on the visual keyboard of the speller. During the experiments, the subjects were asked to stare at the targets in random order. Six blocks of EEG signals were collected from each subject. In each block, there were 40 trials corresponding to all 40 targets. Each trial lasted 6 s in total, including 0.5 s preparation before stimulus onset, 5 s for stimulation, and 0.5 s break after stimulus offset. All data were recorded at 1000 Hz and then down-sampled to 250 Hz.

BETA. The BETA dataset was introduced by Liu et al. [36] in 2020. It includes 64-channel EEG signals from 70 subjects. As shown in Figure 5(b), the speller's targets and frequency range were the same as those in Benchmark, but the keyboard layout was different. A traditional QWERT keyboard layout was used to improve the user experience. There were 4 blocks of EEG signals for each subject, each with 40 trials corresponding to all 40 targets. Each trail still includes 0.5 s preparation before stimulus onset and 0.5 s break after stimulus offset; however, the stimulation duration was reduced to 2 or 3 s.

3.2 Data preprocessing

Russo and Spinelli [37] showed that there is a latency delay in the visual system; hence we extracted EEG signals between [0.14, 1.64] s after each stimulus onset. Nine electrodes (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) around the occipital area were chosen. To remove artifacts and DC drift, EEG signals were bandpass filtered to 7–90 Hz using an eighth-order Butterworth filter.

3.3 Performance measure

Square wave attack is a kind of target attack, so the ASR was used as the performance measure. ASR is the percentage of test samples classified into the target class that the attacker specifies after the perturbations have been added.

3.4 Hyper-parameters

Channels. In a square wave attack, the attacker can choose to attack one or more channels. Obviously, it is easier to attack only one channel. In our experiments, we tried to attack every single channel individually, the four channels (PO3, POz, PO4, and Oz) near the center of the occipital area, and all channels.

Amplitude. When attacking a single channel, the amplitude was determined by its standard deviation. When attacking all channels, the amplitude was determined by the mean standard deviation of all channels. Three different amplitudes (10%/20%/30% of the standard deviation) were used in the experiments.

Phase. In real-world attacks, it is quite difficult for the attacker to know the exact start time of each EEG trial. So, the perturbations should be insensitive to the trial start time. Thus, the phase of the

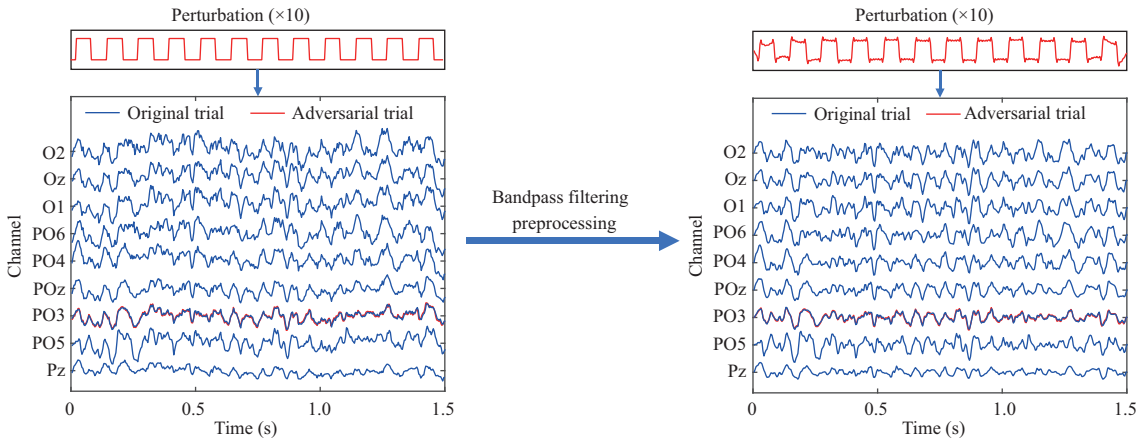


Figure 6 (Color online) An EEG trial from Benchmark before (blue) and after (red) square wave perturbation.

perturbation was randomly chosen for each trial. Experiments showed that the attack performance is insensitive to the perturbation phase, which is good news for practical implementation.

Frequency. In real-world attacks, the frequency of the perturbation depends on the frequency of the target which the attacker wants the system to output. In experiments, we tested our approach on all individual frequencies for each trial.

3.5 Average attack performance on different channels

Table 1 shows the attack results on different channels. Each cell is the average ASR of all subjects in the corresponding dataset. "PO3+POz+PO4+Oz" means attacking the four channels at the same time. "Best of {PO3, POz, PO4, Oz}" is the best ASR of an individual channel in {PO3, POz, PO4, Oz}, which may vary across subjects.

We can get from Table 1 the following.

(1) When a single channel was used, the ASRs of attacking CCA were generally slightly higher than those of FBCCA.

(2) The ASRs varied when different channels were attacked. Particularly, the ASRs of attacking the four individual channels near the center of the occipital area (PO3, POz, PO4, and Oz) were much higher than others, as SSVEPs mainly elicit neural responses in this area.

(3) The ASRs of attacking a single channel were usually higher than those of attacking all channels. When attacking the four channels (PO3, POz, PO4, and Oz) at the same time, the ASRs were highest. However, the best result of attacking one of the four channels in {PO3, POz, PO4, Oz} was only slightly lower than that of attacking the four simultaneously. Since attacking a single channel is easier than attacking four, the former is recommended.

(4) The ASRs on Benchmark were higher than BETA. As introduced in Subsection 3.1, the experimental settings of the two datasets, e.g., the length of each trial, the keyboard layout, were different. Therefore, it is expected that their ASRs are different.

(5) As the amplitude of the perturbation increased, ASRs also increased. When the amplitude was 30% of the standard deviation of the specific channel, the ASR almost reached 100% on Benchmark and 90% on BETA, which is high enough to make a real-world SSVEP speller useless. In a real-world attack, the attacker may attack only one channel, and increase the amplitude of the perturbation gradually to achieve the best trade-off between ASR and stealthiness.

An example of the same EEG trial from Benchmark before and after the attack, without and with preprocessing (band-pass filtering), is shown in Figure 6. The channel under attack was PO3, and the amplitude was 20% of the standard deviation of PO3. The perturbation was too small to be noticed by human eyes. Therefore, it is very difficult to detect the attack by observing the waveform. The perturbation at the top of Figure 6 shows that, although the shape of the square wave perturbation was changed slightly after preprocessing, the main frequency information was retained. So, preprocessing has little influence on the effectiveness of a square wave attack.

Table 1 Average ASRs (%) of square wave attacks on different channels

Channel	Benchmark						BETA					
	CCA			FBCCA			CCA			FBCCA		
	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.
Pz	12.10	43.39	72.99	12.41	45.08	75.44	7.71	23.48	50.80	7.47	23.62	49.11
PO5	44.97	72.63	90.48	42.40	70.27	87.19	24.36	52.21	78.68	23.26	47.20	70.74
PO3	59.59	88.70	**98.36**	57.26	87.11	**97.11**	37.97	69.71	**88.53**	35.70	65.07	**84.26**
POz	34.63	83.97	**97.28**	33.79	82.89	**96.39**	32.62	72.50	**91.69**	29.44	68.48	**87.73**
PO4	45.47	85.07	**98.01**	42.95	83.61	**96.35**	26.41	64.64	**88.85**	24.04	60.04	**84.01**
PO6	34.25	68.31	88.64	31.57	67.13	87.60	9.02	37.25	70.40	8.97	33.99	62.96
O1	30.33	69.01	88.43	27.37	63.68	84.93	25.75	61.02	83.96	23.59	53.09	75.82
Oz	38.09	79.45	93.33	32.87	74.26	90.55	33.47	74.05	90.97	29.28	67.93	85.53
O2	28.66	73.26	88.77	24.70	67.12	86.41	21.02	61.55	85.15	18.96	52.82	77.34
PO3+POz+ PO4+Oz	**78.56**	**98.35**	**99.94**	**77.26**	**98.48**	**99.90**	**61.55**	**94.85**	**99.62**	**56.36**	**90.75**	**98.43**
{PO3, POz, PO4, Oz}	**73.35**	**96.28**	**99.74**	**68.76**	**95.11**	**99.42**	**57.28**	**90.15**	**98.43**	**51.26**	**84.49**	**95.61**
ALL	12.02	40.94	69.79	13.35	47.91	76.53	7.32	20.91	44.58	7.04	21.91	46.16

Table 2 Individual ASRs (%) and ACCs (%) on Benchmark

	CCA				FBCCA			
	Amplitude = 0%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.	Amplitude = 0%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.
Subject	ACC	ASR-PO3/ASR-Best			ACC	ASR-PO3/ASR-Best		
1	52.9	21.47/74.76	86.90/99.68	98.81/100.00	86.3	18.59/62.81	81.94/99.32	98.47/99.98
2	78.3	7.96/97.92	56.08/100.00	89.86/100.00	95.0	7.21/83.78	48.34/99.20	82.20/99.79
3	87.9	69.79/83.09	99.39/99.93	99.98/100.00	98.8	46.03/60.90	89.37/94.73	98.28/99.14
4	93.8	80.59/80.59	99.79/99.94	100.00/100.00	95.0	59.97/59.97	96.90/97.04	99.58/99.63
5	86.7	18.37/18.37	88.98/88.98	99.81/99.81	98.8	22.02/22.02	80.87/80.87	97.83/97.83
6	66.7	41.47/41.47	89.71/92.06	99.23/99.72	93.8	37.61/37.61	85.59/87.16	97.92/98.21
7	55.8	100.00/100.00	100.00/100.00	100.00/100.00	82.5	99.82/99.82	100.00/100.00	100.00/100.00
8	34.6	100.00/100.00	100.00/100.00	100.00/100.00	52.9	99.93/99.93	100.00/100.00	100.00/100.00
9	59.6	5.53/19.99	35.90/85.59	76.45/98.77	81.7	5.25/20.35	36.57/80.56	70.63/97.10
10	80.4	99.56/99.56	100.00/100.00	100.00/100.00	88.3	90.79/90.79	99.93/99.93	100.00/100.00
11	19.2	100.00/100.00	100.00/100.00	100.00/100.00	31.7	100.00/100.00	100.00/100.00	100.00/100.00
12	89.2	49.55/66.42	96.83/99.79	100.00/100.00	98.8	43.87/62.91	85.83/96.57	98.96/99.98
13	68.8	100.00/100.00	100.00/100.00	100.00/100.00	85.0	98.75/98.75	100.00/100.00	100.00/100.00
14	80.4	97.78/97.78	100.00/100.00	100.00/100.00	91.3	94.48/94.48	99.84/99.84	99.98/99.98
15	52.9	83.81/83.81	99.06/99.06	99.78/99.97	68.3	76.47/76.47	98.13/98.13	99.97/99.97
16	20.0	99.85/99.85	100.00/100.00	100.00/100.00	65.8	98.55/98.55	100.00/100.00	100.00/100.00
17	65.8	31.76/43.26	95.74/97.71	99.97/99.98	69.6	44.07/57.02	97.62/98.78	99.93/100.00
18	49.6	30.28/30.28	90.09/91.19	99.13/99.51	74.6	33.71/33.71	89.54/89.54	98.71/98.71
19	25.4	41.85/53.21	88.72/97.38	99.12/99.92	37.9	39.61/47.68	86.10/94.72	98.66/99.92
20	72.9	99.96/99.96	100.00/100.00	100.00/100.00	92.9	97.47/97.47	100.00/100.00	100.00/100.00
21	53.8	97.36/98.48	100.00/100.00	100.00/100.00	75.4	94.44/96.69	99.98/99.98	100.00/100.00
22	84.2	14.54/33.17	81.01/98.10	99.67/99.98	99.2	7.04/18.64	59.64/86.50	94.05/99.56
23	75.0	25.45/60.63	78.39/99.31	98.87/100.00	94.6	20.72/44.59	70.21/93.36	94.71/99.47
24	75.0	10.30/10.95	70.13/78.71	97.73/99.39	94.2	11.98/11.98	71.55/72.53	96.26/96.84
25	92.1	46.79/46.79	87.36/87.36	98.64/98.64	96.7	55.90/55.90	92.50/92.50	99.20/99.20
26	93.3	20.88/79.70	83.81/99.69	99.03/100.00	97.5	43.34/88.66	96.63/99.93	99.79/100.00
27	80.8	97.79/97.79	100.00/100.00	100.00/100.00	97.1	84.78/84.78	99.86/99.86	100.00/100.00
28	63.8	7.66/81.19	59.66/99.79	94.19/100.00	97.5	4.85/66.16	47.70/98.23	86.91/99.93
29	24.6	90.52/90.52	99.96/99.96	100.00/100.00	63.3	87.26/87.26	99.98/99.98	100.00/100.00
30	76.7	99.31/99.31	100.00/100.00	100.00/100.00	86.3	92.48/92.48	99.84/99.84	99.96/99.98
31	85.4	3.36/4.32	41.27/55.49	93.83/95.22	98.8	6.49/9.90	62.86/70.35	93.43/94.63
32	95.0	16.10/98.10	75.79/100.00	98.75/100.00	97.5	18.19/82.01	71.47/99.34	93.51/99.96
33	27.1	82.40/82.40	99.95/99.95	100.00/100.00	50.4	72.60/72.60	99.94/99.94	100.00/100.00
34	88.8	96.84/96.84	100.00/100.00	100.00/100.00	97.5	95.27/95.27	100.00/100.00	100.00/100.00
35	70.4	96.83/96.83	100.00/100.00	100.00/100.00	77.1	94.65/94.65	99.97/99.97	100.00/100.00
Average	66.5	59.59/73.35	88.70/96.28	98.36/99.74	83.2	57.26/68.76	87.11/95.11	97.11/99.42

3.6 Attack performance on individual subjects

Attack results on the individual subjects are shown in Table 2 for Benchmark, and Table 3 for BETA. For each subject, 'ASR-PO3' is the result of attacking channel PO3 only, and 'ASR-Best' is the best result of attacking one of the four channels (PO3, POz, PO4, and Oz).

When the amplitude ratio was 0%, i.e., no attack was applied, the result represented the accuracy (ACC) of CCA/FBCCA on the clean test samples. The ACCs and ASRs of different subjects were different, which is reasonable, because of individual differences among different subjects. As the perturbation amplitude increased, the ASR gradually approached 100%. For some subjects, the ASRs could reach almost 100% with an amplitude ratio of 10%, but for a few other subjects, the amplitude ratio had to be 30% or higher. For most subjects, an amplitude ratio of 20% was enough to make the SSVEP-based BCI useless.

There was no relationship between the ASRs and the data quality of the subject. The ASRs were very high on some subjects whose ACCs were very low, e.g., subjects 11 and 16 in Benchmark. In most cases, the ASRs when attacking channel PO3 were high enough, so PO3 could be the attacker's first choice.

Table 3 Individual ASRs (%) and ACCs (%) on BETA

	CCA				FBCCA			
	Amplitude = 0%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.	Amplitude = 0%std.	Amplitude = 10%std.	Amplitude = 20%std.	Amplitude = 30%std.
Subject	ACC	ASR-PO3/ASR-Best			ACC	ASR-PO3/ASR-Best		
1	88.8	12.01/30.98	74.82/98.74	98.92/99.95	86.3	7.34/15.89	60.71/86.90	94.09/99.57
2	63.8	3.78/8.46	20.75/60.43	56.64/95.78	93.8	3.23/6.78	16.07/57.89	51.93/91.67
3	91.9	100.00/100.00	100.00/100.00	100.00/100.00	92.5	98.46/98.46	100.00/100.00	100.00/100.00
4	48.8	37.95/37.95	88.21/89.29	98.90/99.73	78.8	25.39/25.39	78.37/78.37	96.28/96.31
5	35.0	93.17/96.57	99.89/100.00	100.00/100.00	96.3	88.89/91.62	99.87/100.00	100.00/100.00
6	65.6	100.00/100.00	100.00/100.00	100.00/100.00	75.0	100.00/100.00	100.00/100.00	100.00/100.00
7	32.5	34.79/57.01	94.62/98.32	99.81/100.00	75.6	39.42/55.68	94.99/97.42	99.76/99.82
8	41.9	39.35/74.51	98.32/99.96	100.00/100.00	69.4	52.68/80.17	99.26/99.95	100.00/100.00
9	85.0	5.21/99.56	61.40/100.00	95.57/100.00	86.3	4.07/80.78	39.71/99.67	81.32/99.92
10	57.5	18.81/28.48	76.10/89.79	94.20/97.37	66.9	17.67/22.48	71.54/80.98	91.64/94.79
11	14.4	42.40/42.40	98.54/98.54	100.00/100.00	31.9	33.59/33.59	92.53/92.53	99.81/99.81
12	85.6	3.51/20.89	20.32/72.79	52.89/97.29	92.5	2.98/11.42	11.78/45.25	32.15/77.03
13	76.9	7.25/13.48	55.85/76.85	90.17/97.29	86.3	4.42/6.73	31.59/47.04	69.17/82.39
14	77.5	99.96/99.96	100.00/100.00	100.00/100.00	82.5	99.06/99.79	100.00/100.00	100.00/100.00
15	72.5	19.45/28.82	84.93/95.71	99.51/100.00	87.5	14.70/21.67	61.39/70.42	87.54/92.73
16	72.5	100.00/100.00	100.00/100.00	100.00/100.00	80.6	99.98/100.00	100.00/100.00	100.00/100.00
17	10.0	76.15/94.43	99.89/99.98	100.00/100.00	43.1	88.28/96.81	99.95/100.00	100.00/100.00
18	99.4	58.20/58.20	98.42/98.42	98.76/98.76	98.8	66.57/66.57	96.23/96.23	98.87/98.87
19	65.0	8.85/69.21	80.71/99.96	99.14/100.00	85.6	7.07/46.40	61.93/98.34	94.40/99.98
20	30.6	56.67/56.67	99.62/99.62	100.00/100.00	56.3	68.42/68.42	98.99/98.99	99.96/99.96
21	68.8	4.18/90.34	28.98/100.00	74.09/100.00	93.1	3.31/64.10	21.50/97.98	57.03/99.68
22	71.3	7.43/16.99	60.04/91.92	93.18/99.10	88.1	5.20/10.00	42.89/72.51	84.28/96.81
23	99.4	3.59/3.59	49.31/49.31	97.20/97.20	100.0	3.09/3.09	42.46/42.46	86.65/86.65
24	73.1	3.59/100.00	17.87/100.00	60.04/100.00	81.3	2.89/99.87	12.45/100.00	45.35/100.00
25	81.3	99.95/99.95	100.00/100.00	100.00/100.00	82.5	96.78/96.78	99.78/99.78	100.00/100.00
26	36.3	22.17/22.17	81.25/85.75	97.81/99.24	59.4	14.70/14.70	64.40/64.40	91.64/93.06
27	47.5	3.56/4.87	12.62/25.99	41.26/71.29	85.0	2.98/3.90	7.98/16.14	25.92/47.82
28	71.9	99.40/99.40	99.98/99.98	100.00/100.00	87.5	98.89/98.89	99.68/99.68	99.76/99.76
29	78.1	100.00/100.00	100.00/100.00	100.00/100.00	85.0	99.10/99.10	99.93/99.98	100.00/100.00
30	88.1	13.89/23.35	75.34/91.21	98.50/99.92	83.8	9.46/18.56	62.42/82.68	91.43/97.64
31	21.9	4.67/64.95	26.81/99.79	74.35/100.00	48.1	4.87/46.68	24.78/97.73	63.50/99.98
32	29.4	100.00/100.00	100.00/100.00	100.00/100.00	52.5	99.87/99.87	100.00/100.00	100.00/100.00
33	26.9	100.00/100.00	100.00/100.00	100.00/100.00	41.9	98.92/98.92	100.00/100.00	100.00/100.00
34	61.9	5.10/13.50	34.74/79.89	77.56/98.99	91.9	3.40/6.57	22.60/56.99	65.78/93.85
35	43.8	99.95/100.00	100.00/100.00	100.00/100.00	87.5	99.14/99.68	100.00/100.00	100.00/100.00
36	81.3	3.81/6.98	18.12/44.23	51.75/86.73	94.4	2.78/4.32	8.98/29.43	34.93/69.54
37	93.8	88.81/88.81	100.00/100.00	100.00/100.00	96.9	66.82/66.82	97.20/97.20	99.64/99.64
38	40.6	99.40/99.40	100.00/100.00	100.00/100.00	55.6	96.70/96.70	99.57/99.57	99.68/99.68
39	62.5	10.29/22.18	64.00/82.37	92.32/98.35	76.3	11.75/22.67	63.85/81.12	92.51/97.81
40	56.3	61.65/82.06	99.60/99.98	99.98/100.00	75.6	36.95/51.23	92.17/95.74	99.49/99.79
41	7.5	5.76/10.73	36.26/65.28	81.40/96.17	27.5	5.40/10.84	30.92/59.73	71.34/93.51
42	87.5	27.21/83.18	91.21/100.00	99.85/100.00	88.8	17.64/61.39	77.12/99.18	97.57/99.92
43	58.1	10.46/89.40	65.96/100.00	94.26/100.00	73.8	7.60/68.73	49.20/98.90	84.40/100.00
44	27.5	37.54/83.68	86.32/99.03	98.21/99.81	44.4	37.95/74.87	87.32/96.93	97.74/99.29
45	36.9	6.76/41.32	47.39/98.20	89.90/99.96	70.6	5.42/32.73	35.54/96.01	79.14/100.00
46	35.6	15.07/15.07	86.07/86.07	99.15/99.15	68.1	25.10/25.10	89.39/89.39	99.34/99.34
47	33.1	32.90/72.29	95.78/99.89	99.95/100.00	48.1	32.07/63.92	90.78/99.31	99.46/100.00
48	81.9	89.17/89.17	100.00/100.00	100.00/100.00	92.5	56.31/56.31	96.48/96.48	99.79/99.79
49	93.1	3.20/11.46	16.95/58.34	41.00/94.48	98.8	2.65/6.39	7.37/47.85	26.70/85.49
50	28.8	8.37/14.23	54.54/79.79	91.34/97.48	62.5	12.40/22.03	69.93/82.96	93.71/97.35
51	62.5	42.37/99.98	93.01/100.00	99.37/100.00	76.3	29.53/92.46	82.85/99.95	96.62/100.00
52	76.3	8.01/8.01	62.96/62.96	95.21/95.21	92.5	5.04/5.04	41.89/41.89	80.84/80.84
53	22.5	12.23/89.89	61.98/99.71	90.14/100.00	71.9	16.42/93.23	79.06/99.84	97.03/99.95
54	61.3	23.03/24.59	85.26/86.71	99.07/99.67	64.4	28.09/29.78	90.32/91.26	99.09/99.10
55	20.0	5.90/9.15	33.96/61.00	81.34/95.51	33.8	7.78/14.65	54.20/76.39	89.49/96.64
56	88.1	4.45/35.73	31.12/96.68	75.34/99.93	93.8	3.90/35.12	33.79/93.76	78.35/98.84
57	88.8	3.04/21.42	8.43/33.46	30.01/77.21	92.5	3.04/16.82	12.01/35.62	39.71/68.79
58	71.9	2.65/84.54	3.35/99.98	5.07/100.00	87.5	2.53/43.62	2.70/95.17	3.32/99.54
59	42.5	100.00/100.00	100.00/100.00	100.00/100.00	48.8	100.00/100.00	100.00/100.00	100.00/100.00
60	53.1	100.00/100.00	100.00/100.00	100.00/100.00	78.1	99.62/99.62	100.00/100.00	100.00/100.00
61	14.4	44.37/56.51	97.50/98.74	99.89/99.96	30.0	45.50/56.06	97.15/98.18	99.92/99.98
62	73.1	90.75/90.75	100.00/100.00	100.00/100.00	73.8	76.96/76.96	98.53/98.53	99.85/99.85
63	85.6	7.78/27.04	57.84/94.92	92.15/99.98	95.6	3.18/8.56	27.20/70.98	71.50/96.25
64	21.3	3.57/32.01	21.90/97.45	72.85/99.98	71.3	4.95/50.90	44.42/97.29	82.62/99.89
65	20.6	11.98/58.84	67.46/99.40	95.64/99.96	39.4	12.85/55.78	67.45/98.75	94.39/100.00
66	78.1	15.93/31.29	74.42/90.17	96.25/99.51	90.6	13.25/24.20	58.54/78.64	86.68/94.95
67	96.9	6.43/9.40	59.31/78.70	91.65/99.31	99.4	3.06/3.98	26.96/44.04	66.06/82.17
68	72.5	5.93/24.07	50.54/96.39	94.24/100.00	85.6	3.87/9.78	29.14/61.31	70.59/88.21
69	38.1	100.00/100.00	100.00/100.00	100.00/100.00	78.1	99.67/99.67	100.00/100.00	100.00/100.00
70	77.5	32.25/39.85	97.26/98.54	99.96/100.00	88.1	14.07/18.67	78.98/82.37	98.18/98.70
Average	58.6	37.97/57.28	69.71/90.15	88.53/98.43	75.7	35.70/51.26	65.07/84.49	84.26/95.61

3.7 Parameter sensitivity analysis

This subsection evaluates the sensitivity of three hyper-parameters, i.e., the initial phase of the perturbation, the SSVEP trial length, and the waveform of the perturbation, on Benchmark. The channel under attack was PO3, the amplitude was 20% of its standard deviation, and the target model was CCA.

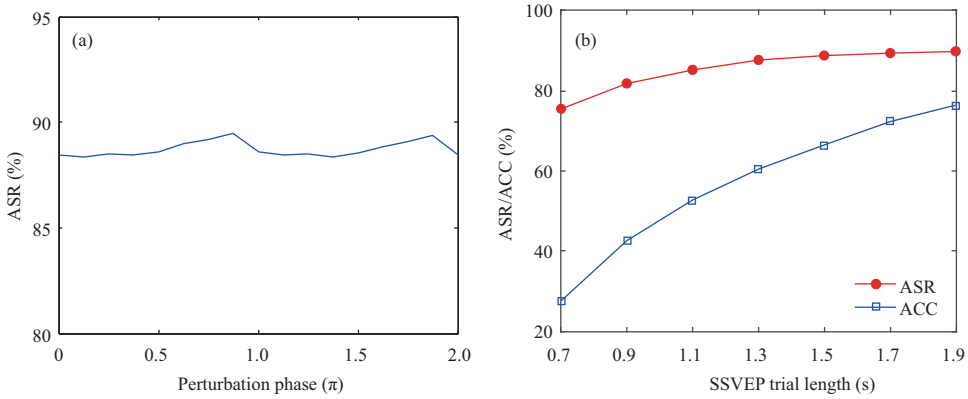


Figure 7 (Color online) ASRs with different perturbation phases (a) and SSVEP trial lengthes (b).

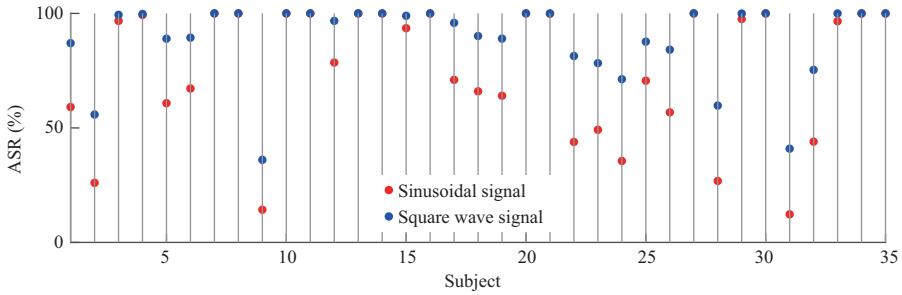


Figure 8 (Color online) ASRs with different waveforms.

As shown in Figure 7(a), when the perturbation phase varied, the ASR only changed slightly, indicating that our attack approach is insensitive to the phase of the perturbation. Thus, the attacker can use any random initial phase.

As shown in Figure 7(b), when the SSVEP trial length increased, the ACC improved significantly, but the ASR changed only slightly; i.e., the impact of the trial length on the ASR was small. In other words, our proposed attack approach is robust to the SSVEP trial length.

We prefer square wave signals to sinusoidal signals as the perturbations, because a square wave signal contains more frequency harmonics than a sinusoidal signal, so a higher ASR can be obtained. To demonstrate this, we show their respective experimental results in Figure 8. For all subjects, square wave signals achieved higher ASRs than, or the same ASRs as, sinusoidal signals.

4 Conclusion and future research

Existing evasion attacks to EEG-based BCIs [23, 30, 31] did not consider the causality and difficulty constraints of implementing them in a real-world BCI system. Inspired by the frequency characteristics of SSVEP, this paper proposes to use square wave signals for target attacks. Due to the simplicity of square wave signals, the attack is easy to implement in real-world SSVEP-based BCI systems. Experiments showed that our proposed square wave attack approach has high ASR and is difficult to detect. It becomes a significant threat to SSVEP-based BCIs.

The main limitation of our approach is that only training-free models like CCA and FBCCA were considered. However, multiple novel training-based approaches have been proposed recently, e.g., extended CCA [34], multi-stimulus TRCA [38], (e)TRCA-Tu [39]. They are also gaining popularity and demonstrating promising performance in SSVEP classification. We have tried to attack these classifiers, but a much larger perturbation amplitude was needed, making the perturbations easier to detect. More significantly, the square wave perturbation needs to be added at the start of each EEG trial, which is difficult to implement. Our future research will tackle these problems for training-based classifiers in SSVEP-based BCIs.

Additionally, we will also demonstrate square wave attacks on a real-world SSVEP-based BCI system. But most importantly, we will investigate approaches to defend against square wave attacks, because the

ultimate goal of our research is to discover and eliminate security risks of EEG-based BCIs, instead of damaging them.

Acknowledgements This work was supported by Open Research Projects of Zhejiang Lab (Grant No. 2021KE0AB04) and Technology Innovation Project of Hubei Province of China (Grant No. 2019AEA171).

References

1 Graimann B, Allison B, Pfurtscheller G. Brain-computer interfaces: a gentle introduction. In: Brain-Computer Interfaces. Berlin: Springer, 2010. 1–27
2 Nicolas-Alonso L F, Gomez-Gil J. Brain computer interfaces, a review. Sensors, 2012, 12: 1211–1279
3 Sutton S, Braren M, Zubin J, et al. Evoked-potential correlates of stimulus uncertainty. Science, 1965, 150: 1187–1188
4 Wu D R, Lawhern V J, Hairston W D, et al. Switching EEG headsets made easy: reducing offline calibration effort using active weighted adaptation regularization. IEEE Trans Neural Syst Rehabil Eng, 2016, 24: 1125–1137
5 Wu D R. Online and offline domain adaptation for reducing BCI calibration effort. IEEE Trans Human-Mach Syst, 2017, 47: 550–563
6 Farwell L A, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography Clin Neurophysiol, 1988, 70: 510–523
7 Yu Y, Liu Y D, Yin E W, et al. An asynchronous hybrid spelling approach based on EEG-EOG signals for Chinese character input. IEEE Trans Neural Syst Rehabil Eng, 2019, 27: 1292–1302
8 Jin J, Xiao R C, Daly I, et al. Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory. IEEE Trans Neural Netw Learn Syst, 2021, 32: 4814–4825
9 Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proc IEEE, 2001, 89: 1123–1134
10 Zhu D H, Bieger J, Molina G G, et al. A survey of stimulation methods used in SSVEP-based BCIs. Comput Intell Neurosci, 2010, 2010: 1–12
11 Yin E, Zhou Z, Jiang J, et al. A dynamically optimized SSVEP brain-computer interface (BCI) speller. IEEE Trans Biomed Eng, 2015, 62: 1447–1456
12 Vialatte F B, Maurice M, Dauwels J, et al. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Prog Neurobiol, 2010, 90: 418–438
13 Beverina F, Palmas G, Silvoni S, et al. User adaptive BCIs: SSVEP and P300 based interfaces. Psychnol J, 2003, 1: 331–354
14 Zhang N N, Liu Y D, Yin E W, et al. Retinotopic and topographic analyses with gaze restriction for steady-state visual evoked potentials. Sci Rep, 2019, 9: 4472
15 Zhang Y S, Yin E W, Li F L, et al. Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs. Neural Networks, 2019, 119: 1–9
16 Lin Z L, Zhang C S, Wu W, et al. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans Biomed Eng, 2006, 53: 2610–2614
17 Chen X G, Wang Y J, Gao S K, et al. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. J Neural Eng, 2015, 12: 046008
18 Nakanishi M, Wang Y J, Chen X G, et al. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. IEEE Trans Biomed Eng, 2018, 65: 104–112
19 Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. In: Proceedings of International Conference on Learning Representations, Banff, 2014
20 Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: Proceedings of International Conference on Learning Representations, San Diego, 2015
21 Carlini N, Wagner D. Audio adversarial examples: targeted attacks on speech-to-text. In: Proceedings of IEEE Symposium on Security and Privacy, San Francisco, 2018
22 Han X T, Hu Y X, Foschini L, et al. Deep learning models for electrocardiograms are susceptible to adversarial attack. Nat Med, 2020, 26: 360–363
23 Zhang X, Wu D R. On the vulnerability of CNN classifiers in EEG-based BCIs. IEEE Trans Neural Syst Rehabil Eng, 2019, 27: 814–825
24 Moosavi-Dezfooli S M, Fawzi A, Frossard P. DeepFool: a simple and accurate method to fool deep neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 2574–2582
25 Madry A, Makelov A, Schmidt A, et al. Towards deep learning models resistant to adversarial attacks. In: Proceedings of International Conference on Learning Representations, Vancouver, 2018
26 Xiao H, Biggio B, Brown G, et al. Is feature selection secure against training data poisoning? In: Proceedings of International Conference on Machine Learning, Lille, 2015. 1689–1698
27 Muñoz-González L, Biggio B, Demontis A, et al. Towards poisoning of deep learning algorithms with back-gradient optimization. In: Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, Dallas, 2017. 27–38
28 Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. J Neural Eng, 2018, 15: 056013

29 Schirrmeister R T, Springenberg J T, Fiederer L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp, 2017, 38: 5391–5420

30 Liu Z H, Meng L B, Zhang X, et al. Universal adversarial perturbations for CNN classifiers in EEG-based BCIs. J Neural Eng, 2021, 18: 0460a4

31 Zhang X, Wu D R, Ding L Y, et al. Tiny noise, big mistakes: adversarial perturbations induce errors in brain-computer interface spellers. Natl Sci Rev, 2021, 8: 4

32 Meng L B, Huang J, Zeng Z G, et al. EEG-based brain-computer interfaces are vulnerable to backdoor attacks. 2021. ArXiv:2011.00101

33 Chen X G, Chen Z K, Gao S K, et al. A high-ITR SSVEP-based BCI speller. Brain-Comput Interface, 2014, 1: 181–191

34 Nakanishi M, Wang Y J, Wang Y T, et al. Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate. PLOS One, 2014, 9: e99235

35 Wang Y J, Chen X G, Gao X R, et al. A benchmark dataset for SSVEP-based brain-computer interfaces. IEEE Trans Neural Syst Rehabil Eng, 2017, 25: 1746–1752

36 Liu B C, Huang X S, Wang Y J, et al. BETA: a large benchmark database toward SSVEP-BCI application. Front Neurosci, 2020, 14: 627

37 Russo F D, Spinelli D. Electrophysiological evidence for an early attentional mechanism in visual processing in humans. Vision Res, 1999, 39: 2975–2985

38 Wong C M, Wan F, Wang B, et al. Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs. J Neural Eng, 2020, 17: 016026

39 Jin J, Wang Z Q, Xu R, et al. Robust similarity measurement based on a novel time filter for SSVEPs detection. IEEE Trans Neural Netw Learn Syst, 2021. doi: 10.1109/TNNLS.2021.3118468