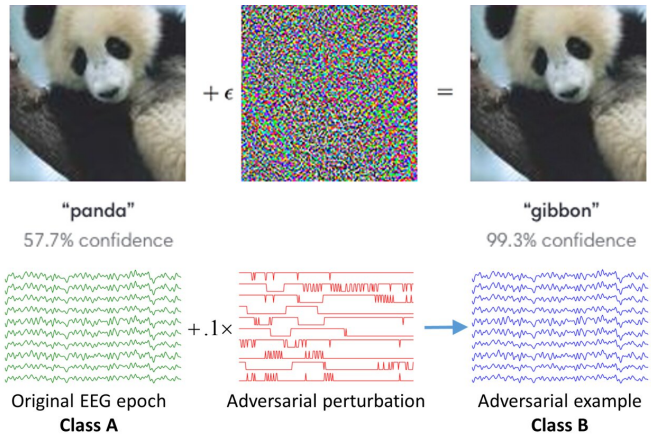
Study unveils security vulnerabilities in EEG-based brain-computer interfaces

13 February 2020, by Ingrid Fadelli



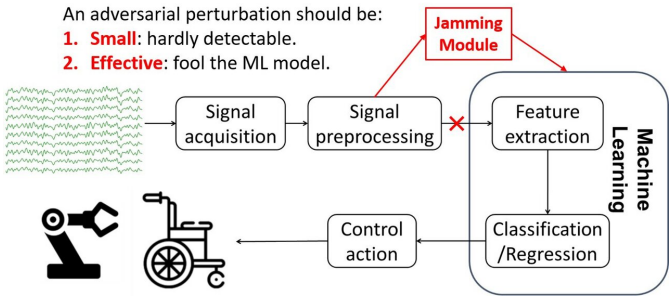
Credit: Zhang et al.

Brain-computer interfaces (BCIs) are tools that can connect the human brain with an electronic device, typically using electroencephalography (EEG). In recent years, advances in machine learning (ML) have enabled the development of more advanced BCI spellers, devices that allow people to communicate with computers using their thoughts.

So far, most studies in this area have focused on developing BCI classifiers that are faster and more reliable, rather than investigating their possible security vulnerabilities. Recent research, however, suggests that machine learning algorithms can sometimes be fooled by attackers, whether they are used in computer vision, speech recognition, or other domains. This is often done using adversarial examples, which are tiny perturbations in data that are indistinguishable by humans.

Researchers at Huazhong University of Science and Technology have recently carried out a study investigating the security of EEG-based BCI spellers, and more specifically, how they are affected by adversarial perturbations. Their paper, pre-published on arXiv, suggests that BCI spellers are fooled by these perturbations and are thus highly vulnerable to adversarial attacks.

"This article aims to expose a critical security concern in EEG-based BCI spellers and more broadly, EEG-based BCIs, which has received little attention before," Dongrui Wu, one of the researchers who carried out the study, told TechXplore. "It shows that one can generate tiny adversarial EEG perturbation templates for target attacks for both P300 and steady-state visual evoked potential (SSVEP) spellers, i.e., mislead the classification to any character the attacker wants, regardless of what the user-intended character is."



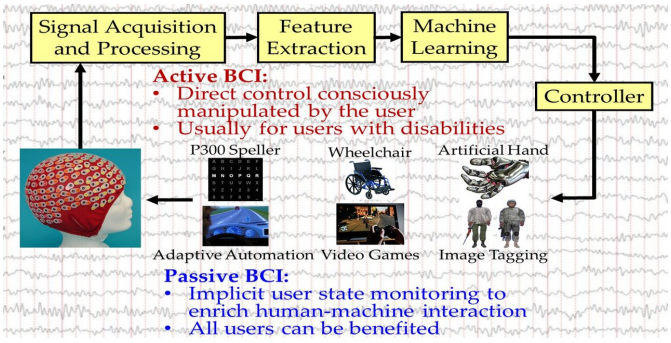
Credit: Zhang et al.

P300 BCI spellers are already used in several settings, including in clinics, to evaluate or detect disorders of consciousness. Adversarial attacks to BCI spellers could thus have numerous consequences, ranging from simple usability issues to severe patient misdiagnoses.

"We believe that a new and more detailed understanding of how adversarial EEG perturbations affect BCI classification can inform the design of BCIs to defend against such attacks," Wu explained.

Wu and his colleagues found that to carry out a successful adversarial attack on a BCI speller, the attacker only needs to access some of the data used to train the device. He/she can use this data to train the perturbation template, subsequently adding the template to benign EEG trials to perform the attack.

Current approaches for conducting adversarial attacks have two main limitations. First, they require some subject-specific EEG samples to create the adversarial perturbation template. Second, to perform the attack more effectively, the attacker needs to know the exact timing of the EEG stimulus. If the attacker successfully overcomes these limitations, the impact of his/her attack could be far greater.

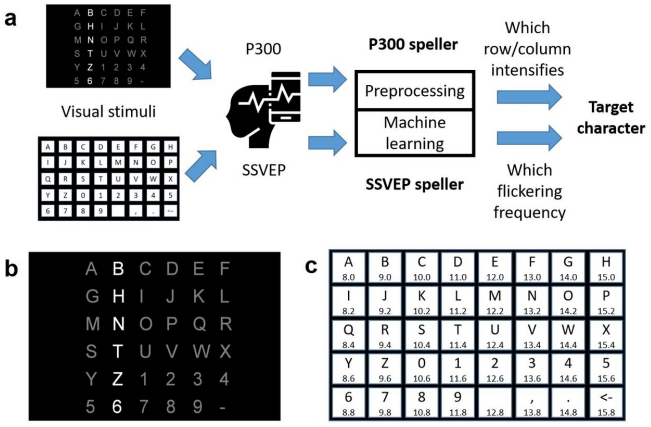


Credit: Zhang et al.

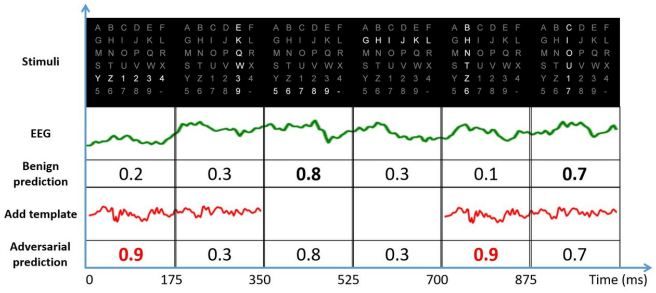
"Defending adversarial attacks is a common research problem in many applications of machine learning, e.g., computer vision, speech recognition, and BCIs," Wu said. "If we know the approach the attacker takes, then we can develop strategies to defend against it, just like how we defend against computer viruses: a virus breaks out first and then we find ways to kill it."

Attackers are always trying to come up with new ways to circumvent security measures, so it is important for researchers to continue investigating system vulnerabilities and come up with new security measures. While it is unavoidable for targeted security solutions to be developed after a specific vulnerability has been identified, uncovering general issues with a system and taking precautions can be very useful.

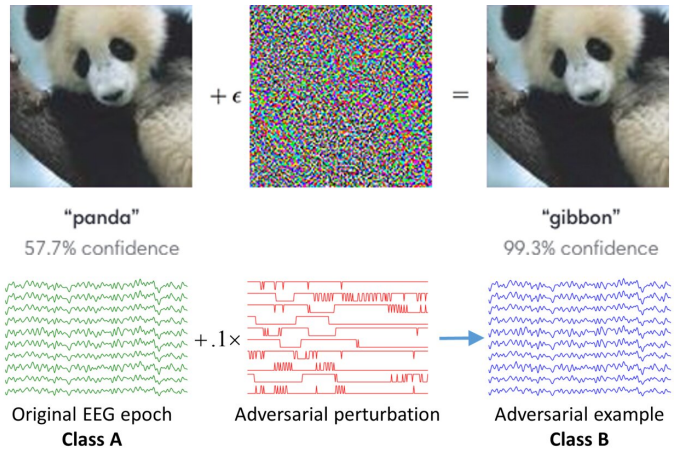
The study carried out by Wu and his colleagues has helped to unveil general security risks associated with EEG-based BCIs. Their findings could help to identify tentative solutions that could decrease the impact of adversarial attacks on these devices.



Credit: Zhang et al.



Credit: Zhang et al.



Credit: Zhang et al.

Wu and his colleagues hope that their research will encourage others to investigate the limitations and vulnerabilities of EEG-based spellers or other BCI devices. Their findings could ultimately pave the way towards the development of techniques to strengthen the security of BCIs, preventing misdiagnoses and other undesired effects of adversarial attacks.

"We want to emphasize that the goal of this study is not to damage EEG-based BCIs, but to demonstrate that serious adversarial attacks to EEG-based BCIs are possible and hence expose a critical security concern that received little attention before," Wu said. "In our future research, we plan to develop strategies to defend against such attacks. Meanwhile, we hope that our study can attract more researchers' attention to the security of EEG-based BCIs."

More information: Tiny noise can make an EEG-based brain-computer interface speller output anything. arXiv:2001.11569 [cs.HC]. arxiv.org/abs/2001.11569