On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces

Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito,
Tomas Ros, Dawn Song

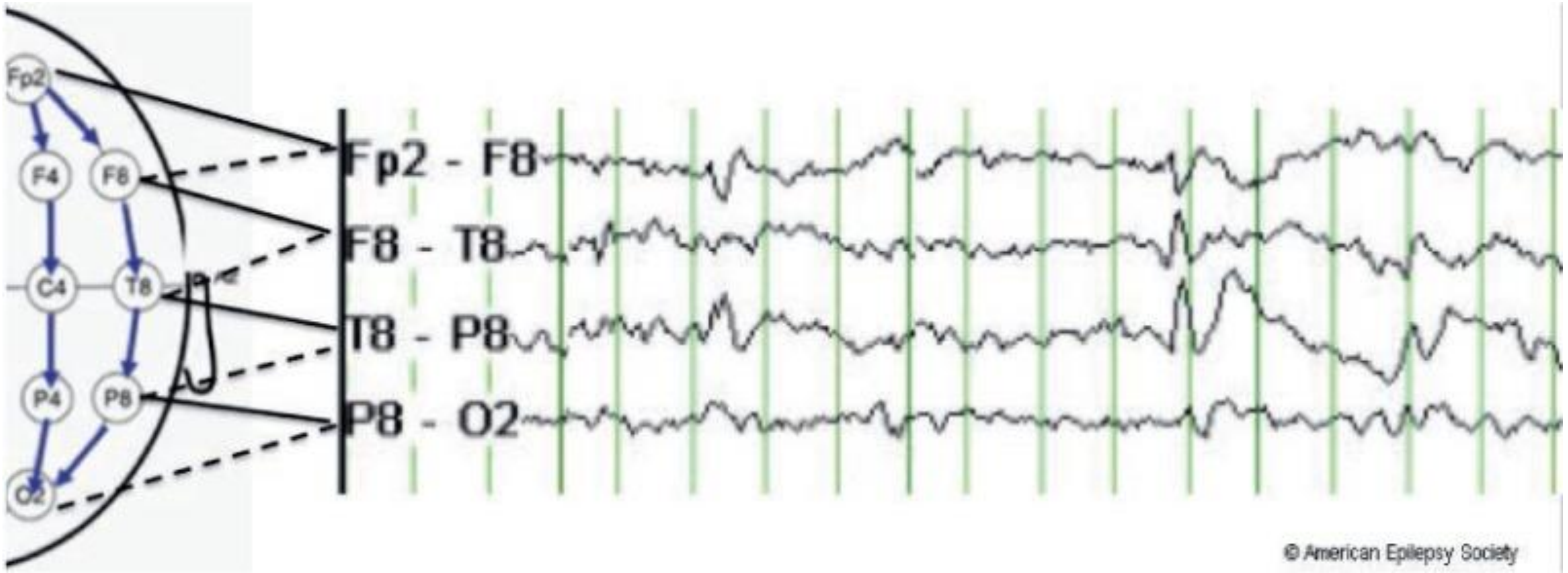
Presenter: Hwijoon Lim

Brain-Computer Interfaces (BCIs)

- BCIs enable a non-muscular communication between a user and an external device

Electroencephalography (EEG)

- Non-invasive method
- The EEG signal is recorded from scalp electrodes and continuously sampled (typically 128Hz – 512Hz)



BCI devices

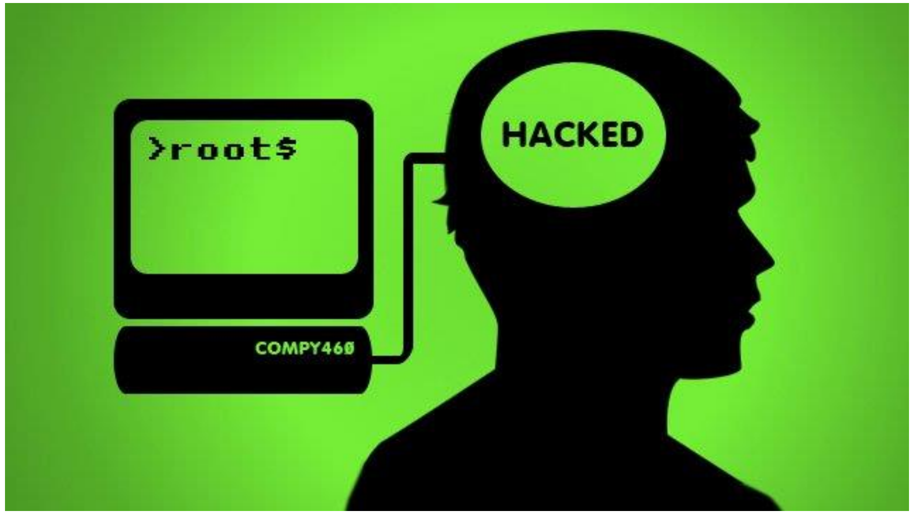
- Consumer-grade BCI devices
  - Low-cost EEG-based gaming devices are offered by Emotiv Systems and NeuroSky

An EPOC device (Emotiv Systems)

A MindSet device (NeuroSky)

BCI devices

- Consumer-grade BCI devices
  - Software development kits to support the expansion of tools and games
  - Such as a mind-controlled keyboard and mouse and hands-free arcade games
- How third-party EEG applications could infer private information about the user?

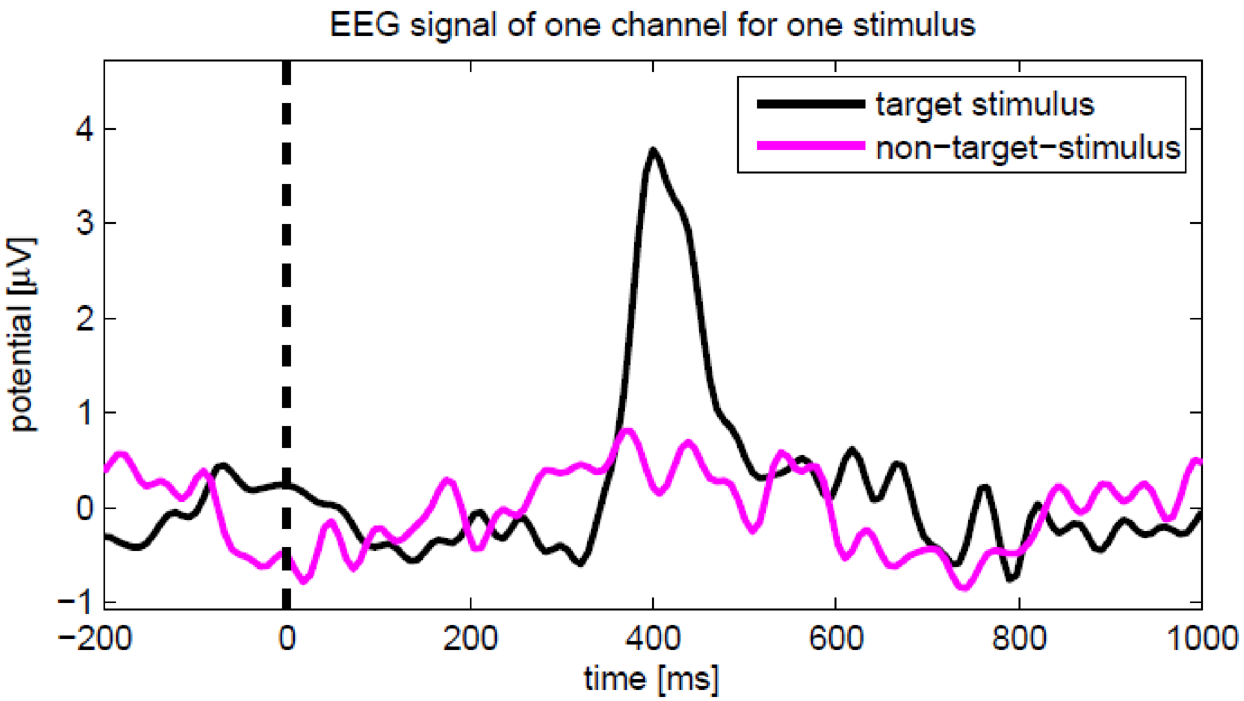


Event-Related Potential (ERP)

- An ERP is detected as a pattern of voltage change after a certain auditory or visual stimulus is presented
- Every ERP is time-locked to the stimulus
- The most prominent ERP component is the P300

P300

- P300 can be detected as an amplitude peak in the EEG signal at about 300ms after the stimulus



Related Works

- EEG-based identification (Poulos et al., 1999)
  - It achieved a high true positive rate and a high true negative rate
- EEG-based authentication (Marcel et al, 2007)
  - Instead of typing a password, it requires the user to think of password
- Key generation technique resistant against coercion attacks (Gupta et al, 2010)
  - Incorporate the user’s emotional status through skin conductance measurements
- Assisting a user in efficient search (Van Vliet et al, 2010)
  - An ERP called N400
- Guilty-Knowledge Test (Abootalebi et al, 2009)
  - Use P300 in lie detection

BCI Application

- BCI devices have “App Stores” like an application stores for smart phones
  - The applications are developed by third parties
  - Provide unrestricted access to the raw EEG signal
  - Applications can control the contents for users

Threat Model

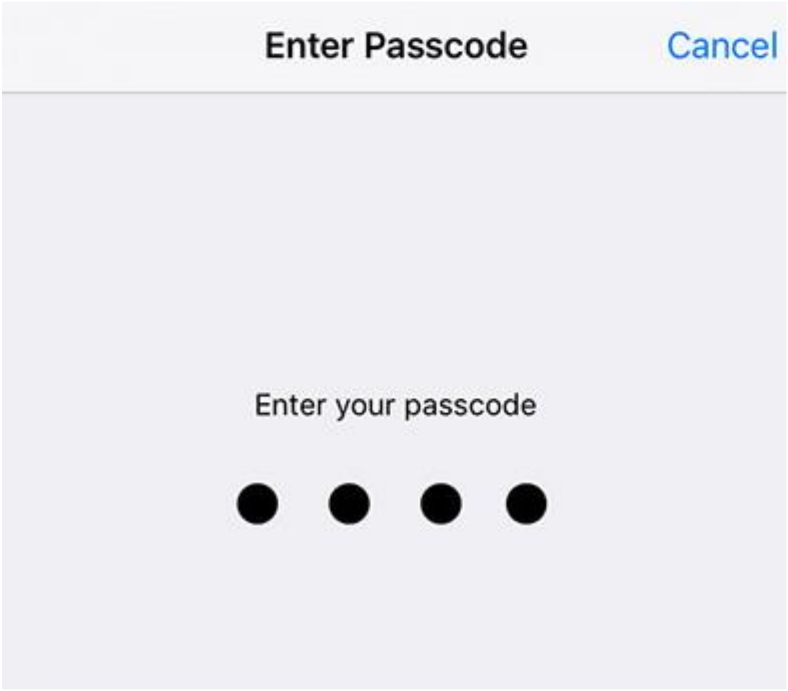
- The attacker is a malicious third-party developer
- His goal is to learn as much information as possible about the user without any malware
- The attacker can read the EEG signal from the device and can display text, videos and images on the screen

Experiment

- Each experiment consisted of three main steps:
  1. (optional) Brief verbal explanation of the task by the operator
  2. (optional) Message on screen for 2 seconds
  3. Images being flashed in random order for the duration of the experiment
- Total 5 Experiments

Experiment 1

- Pin Code
  - Choose and memorize random PIN!
  - Enter the first digit of PIN at the end of the experiment

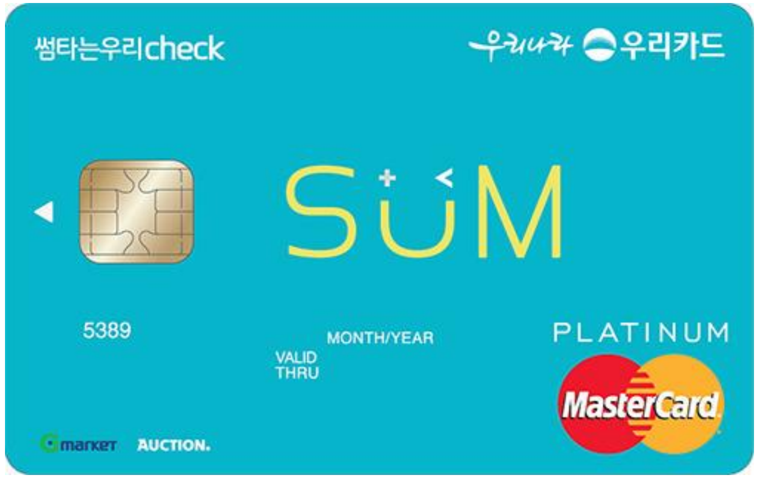


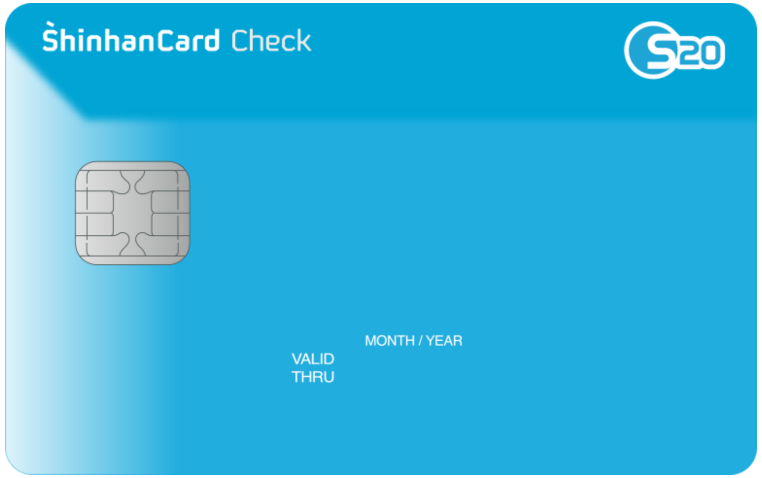
Experiment 2

- Bank Information
  - Just show the logo of different banks

Experiment 2

- Bank Information
  - Show the images of the debit card





Experiment 3

- Month of Birth
  - On-screen message: Which month were you born?
  - Flashing the name of the months randomly

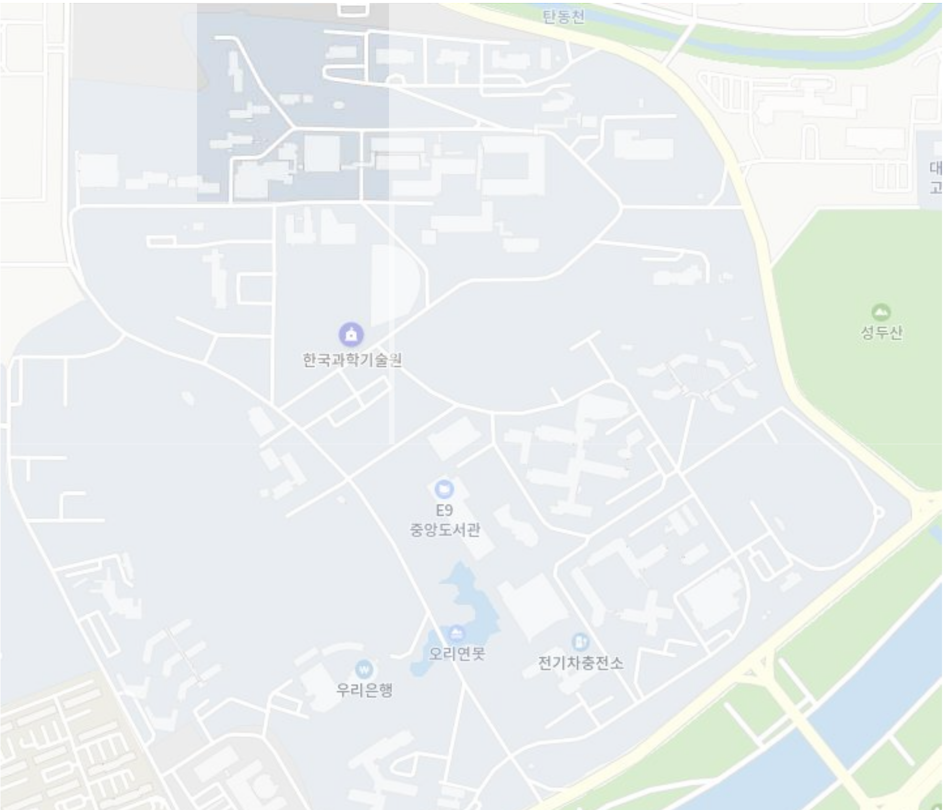
December

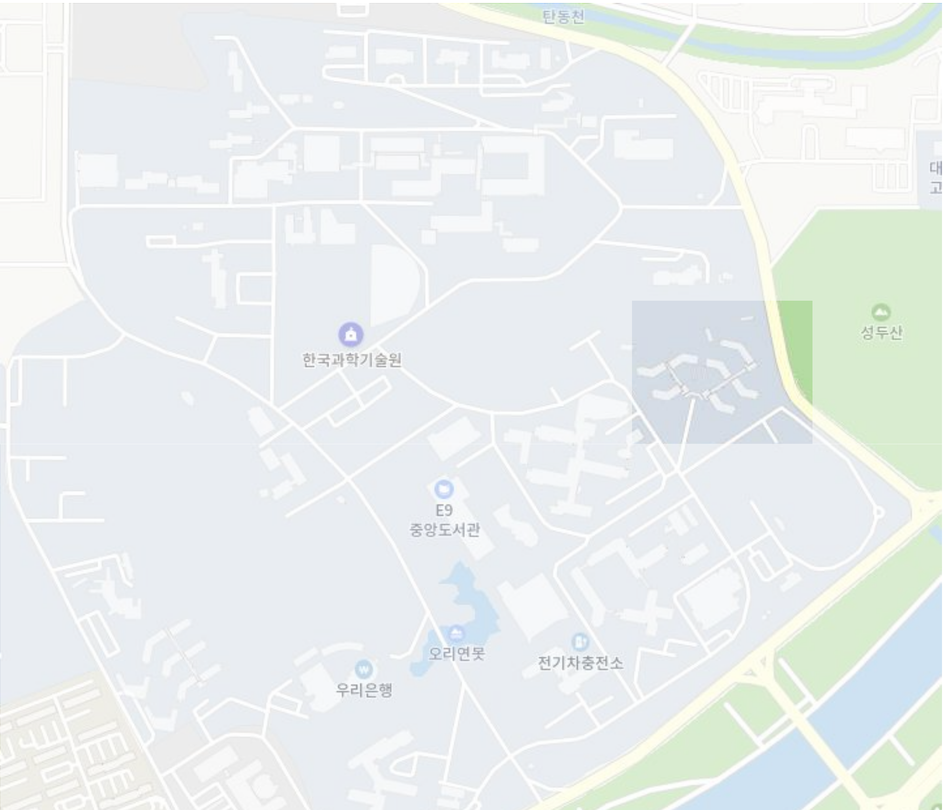
Experiment 4

- Face Recognition
  - On-screen message: Do you know any of these people?

Experiment 5

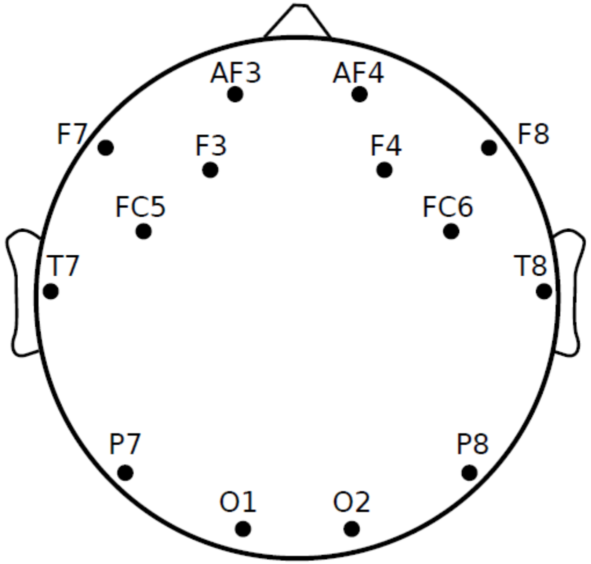
- Geographic Location
  - Show highlighted maps with different highlighted zone





Data Collection

- The amplitudes of the EEG signal are recoded with 14 different electrodes
- The gaming device is not made for detecting P300
  - The P300 is mostly detected at the parietal lobe
  - But, they have more electrodes on the frontal part of the scalp



Data Collection

- Each channel is recorded at a sampling rate of 128Hz
- After showing stimuli to the user, output the time stamp and the indicator of each stimulus
- Obtain the tuple of (EEG signal, the stimuli)

Data Collection – Challenges

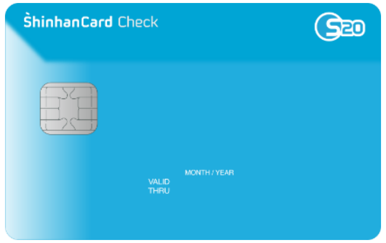
- The attack vector exploits the occurrence of P300 peaks
  - The attack vector must …
    - detect P300 peaks reliably
    - discriminate peaks from all other EEG signals measured on non-target stimuli

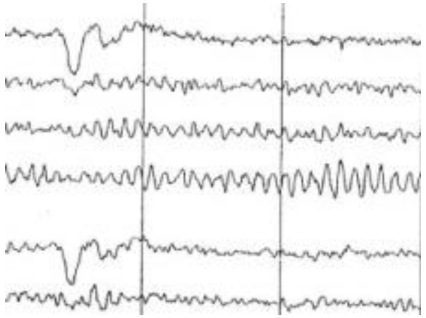
- The user do not intend to provide a discriminative signal for the target stimuli

→ **Train classifier to detect P300 peaks and corresponding stimuli!**

Classification of Target Stimuli

- Input : EEG data (called epochs) associated with a stimulus
  - Each EEG data starts a few milliseconds prior to the stimulus
  - Each EEG data ends 800ms - 1500ms after the stimulus

(

, ) (

, )

- The classification task = Training phase + Classification phase

Training Phase

- Train how to tell if input epoch is generated is target stimulus

- Input
  - A set of epochs $x \in X^{train}$
  - A vector of label $y \in Y$
- Output
  - A function g that maps epochs to target stimuli labels:
  - $g(x) = y$

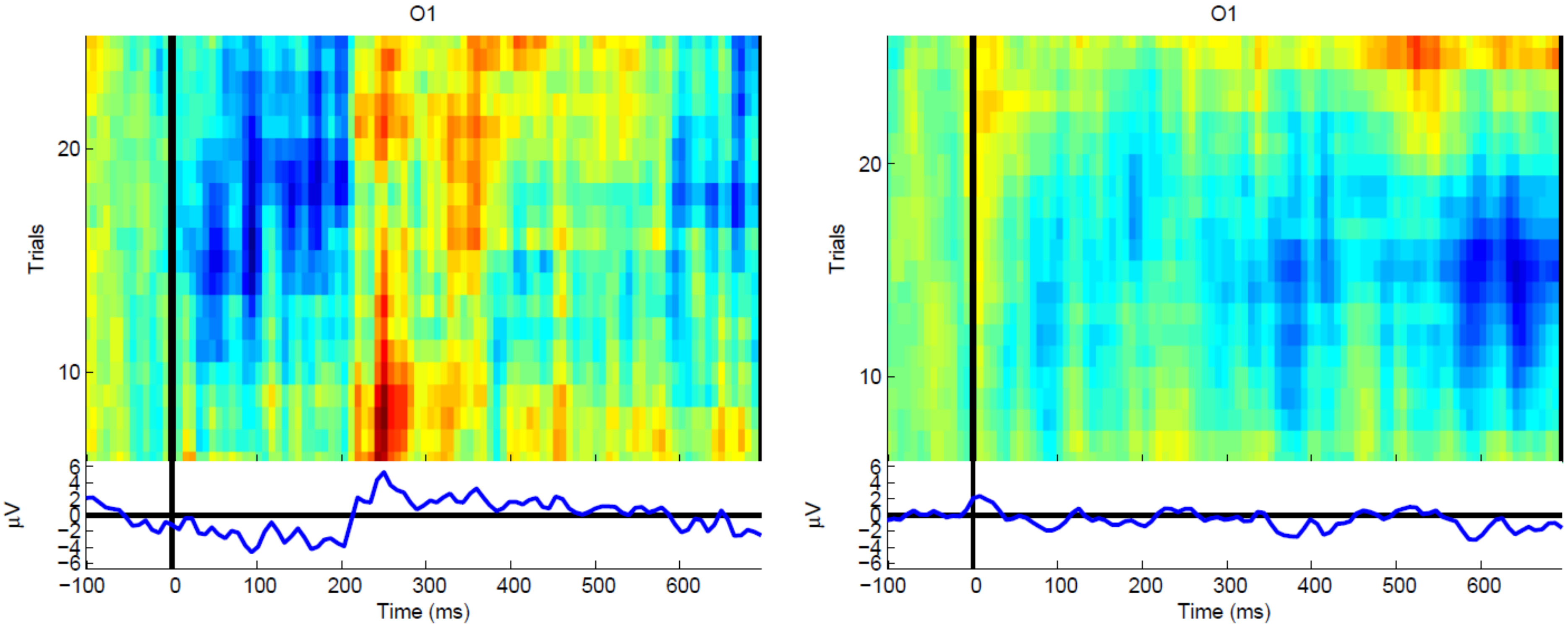
Classification Phase

- Use model from training phase to obtain stimulus from given epoch

- Input
  - A set of new epochs $x^{test} \in X^{test}$
- Output
  - A set of estimation $\{\hat{y} = g(x^{test})\}$

Classification phase

- For stimulus k, $N_k^{(+)}$ is the sum of y's that are associated with stimulus k



(a) target stimulus (b) non-target stimulus

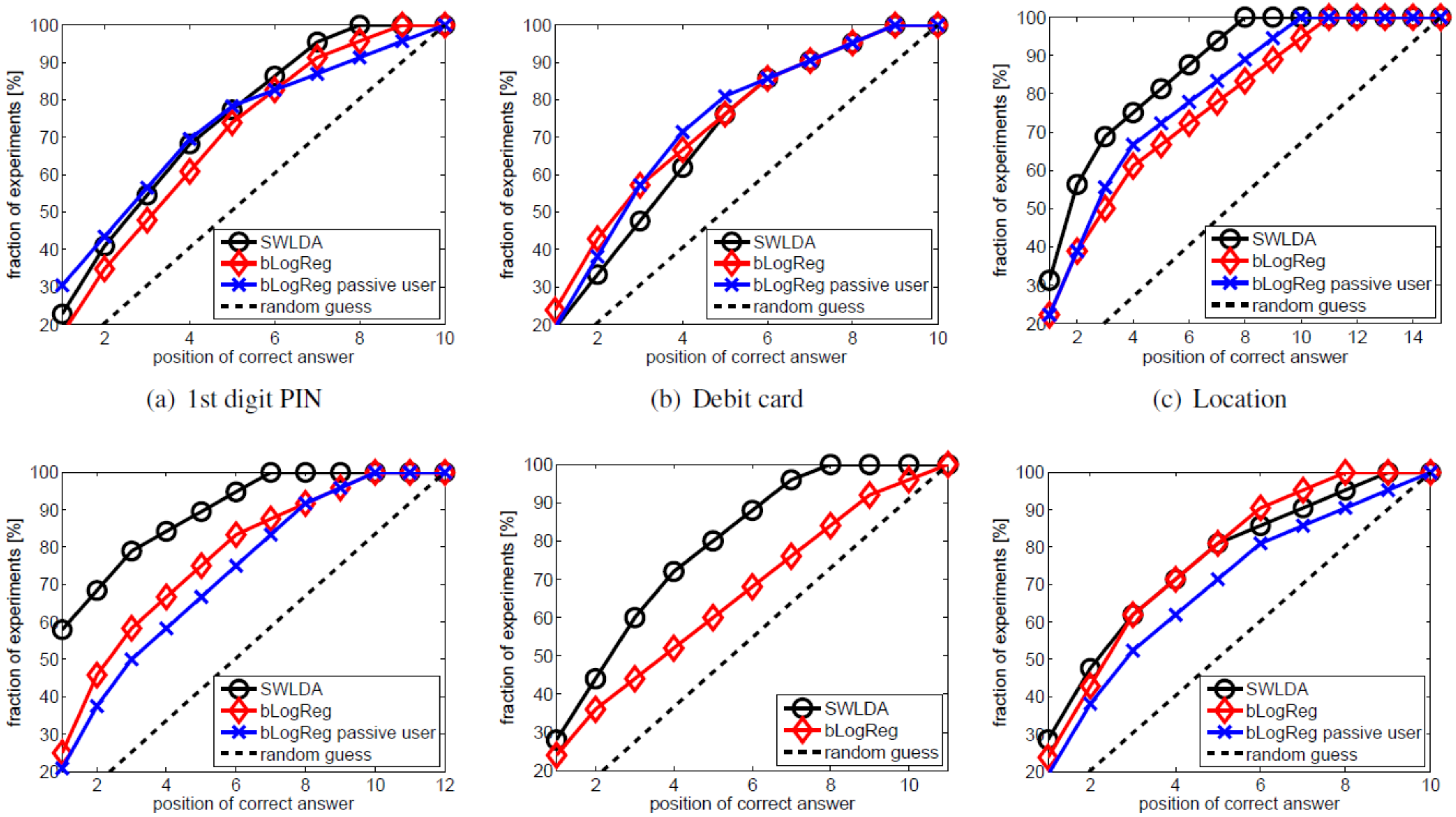
The Classifier Function

- Boosted logistic regression (bLogReg)
  - The model is trained on the training data by minimizing the negative Bernoulli log-likelihood of the model
- Stepwise Linear Discriminant Analysis (SWLDA)
  - Extension of Fisher's linear discriminant analysis (LDA)
  - More robust to noise

Two Training Situation

- User-supported calibration
  - Actively support the training phase
  - Do not support the detection with new stimuli
- On-the fly calibration
  - Do not support the training phase
  - Do not support the detection with new stimuli

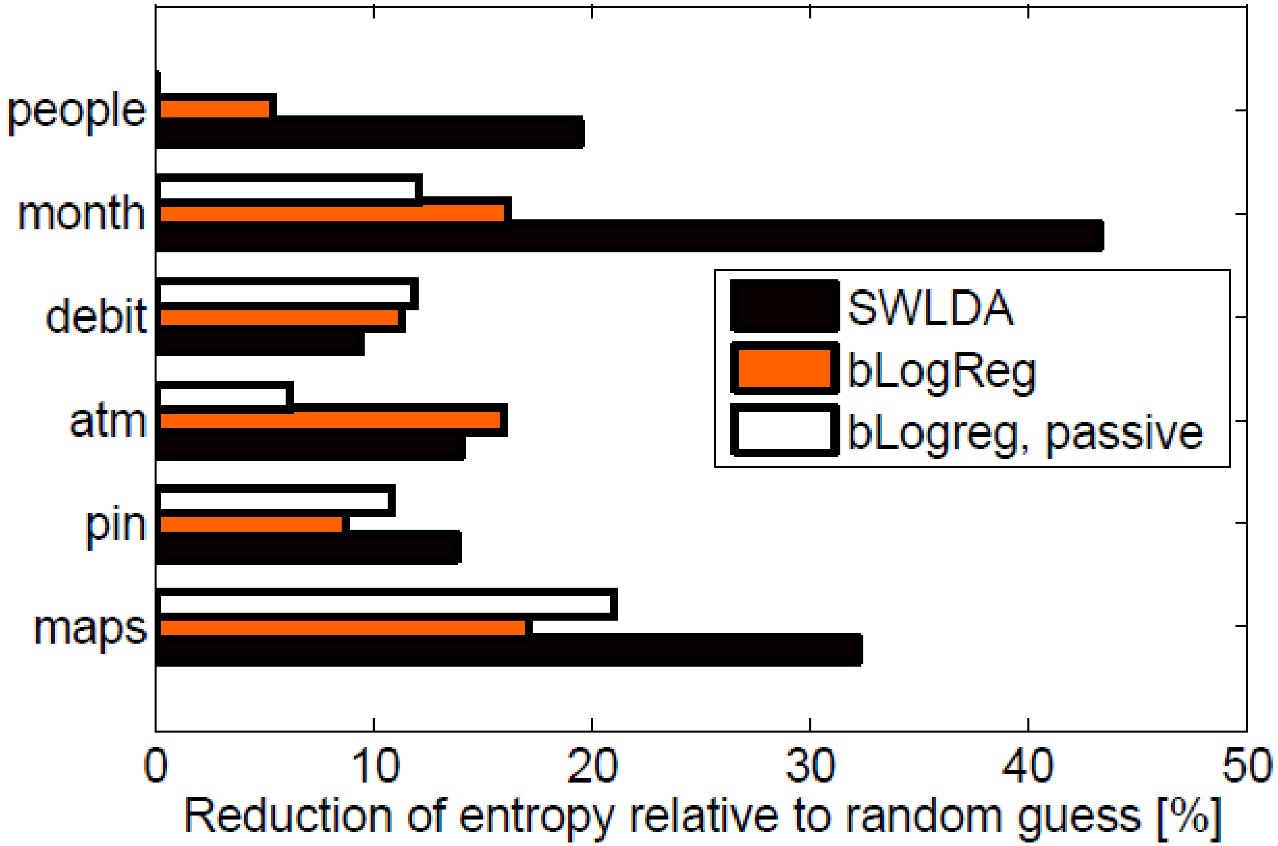
Result



(a) 1st digit PIN (b) Debit card (c) Location

(d) Month of birth (e) People (f) ATM machine

Result


people
month
debit
atm
pin
maps
SWLDA
bLogReg
bLogreg, passive
0
10
20
30
40
50
Reduction of entropy relative to random guess [%]

Defense

- Users of the BCI devices could actively try to hinder probing
  - Concentrate on non-target stimuli
  - Not realistic
- Do not expose the raw data from EEG devices to third-party applications
  - The EEG vendor would create a restricted API
- Add noise to the EEG raw data
  - It could decrease accuracy of legitimate applications

Future works

- "Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity" (Marcello et al, 2016)
  - Research on "Brain-hacking"
- "Side-Channel Attacks Against the Human Brain: the PIN Code Case Study" (Lange et al, 2017)
  - Extract concrete PIN codes from EEG signals
- "Detection of Subconscious Face Recognition Using Consumer-Grade Brain-Computer Interfaces" (Martin et al, 2016)
  - Study subconscious face recognition