Emotiv Epoc+ live metrics quality validation

Yuan Lyu
Human Computer Interaction and Design
384573
yuan.lyu@campus.tu-berlin.de

Stéphane Garti
Interaction Design
Business Management
387491
stgarti@campus.tu-berlin.de

Mikko Honkanen
Human Computer Interaction and Design
387608
mikko.honkanen@campus.tu-berlin.de

ABSTRACT

In the last 10 years, many commercial EEG devices have appeared on the market. The arrival of these low-cost sensors has been accompanied by marketing efforts to promote EEG devices as controllers for games, training systems for the brain, or as introspection methods into mind and emotions. Compared to the traditional laboratory grade EEG measurement tools, these low cost alternatives provide a turnkey solution for the live analysis of the data. One of these consumer grade devices is the Emotiv Epoc+, that provides among other live metrics of the users' state, the level of arousal of the users. In this study we evaluate how well the arousal level provided by the Emotiv Xavier SDK correlates with the participants subjective view.

While measuring their arousal level with the Emotiv headset, participants were shown pictures for which they had to rate their reaction using the Affective Slider. The results showed an average linear correlation inferior to 0.3 for the arousal level acquired through the Emotiv headset and the arousal level from the Affective slider. Hence, no correlation was found between the arousal level from the Emotiv headset and the Affective Slider. Some attempts to clean the data have been tried out, with no significant results on the correlation rate. We conclude that the Emotiv headset live metrics should not be used in critical application like research, but can still be suitable for less critical applications, like for example gaming.

1. INTRODUCTION

Research of Brain-Computer Interfaces has traditionally been done in lab environment due to the expensive and sensible equipment needed. Low cost alternatives to the laboratory level equipment that enable the use of EEG in out-of-the-lab applications have come to the market. [1] Emotiv Epoc+ is one of these low cost EEG sensors, that are easy to use and provide real time EEG data analysis out of the box. It provides a turnkey software solution to display user's level of "Engagement / Boredom", "Frustration", "Meditation", "Short Term Excitement" in real-time, called Live Metrics. [2]

The quality of the raw data provided by the Emotiv Epoc+ has been reviewed in several studies [1] [3] [4]. The main results are that the data quality allows the recognition of basic artefacts, but that it should not be used for critical applications. However, the relevancy of the Live Metrics provided by the software Emotiv Xavier SDK has not been reviewed by independent studies. As the EEG technology provides a lot of opportunities for applications, it would be beneficial to know, how well the real time analysis corresponds to the emotions of the user. Therefore it would be beneficial to study the quality of the real time emotion tracking provided by the Emotiv Xavier SDK.

The raw data provided by the Emotiv has been studied and compared with laboratory level devices [1] [3] [4]. Duvinage et. at (2013) conclude that the Emotiv headset does not perform as well as a medical device and that the relative operational and maintenance costs are higher than those of the medical-grade competitors [1]. Badcock et. al (2013) state that Emotiv may prove a valid alternative to medical grade systems when recording late auditory ERPs. They also state that the system may also be useful in the future for measuring less reliable ERPs, if their detection can be made as reliable as of the auditory ERPs. [3] Stytsenko et. al (2011) suggest that the data between the Emotiv headset and a G-TEC device are alike in general, but a G-TEC device has a cleaner and stronger signal. In addition, they evidenced a drift in

the recording speed of the headset. [4]

To show any evidence of relevancy of the processed data that comes from Emotiv Software, we need to have a reference to compare with. One of the widely used methods is the Self Assessment Manikin (SAM) [5]. The Self Assessment Manikin uses pictures to directly measure the pleasure, arousal, and dominance levels that are associated with a user's affective reaction to a certain stimuli [5]. Each of the three states are rated by the user on a scale from one to five with the help of a representative picture [5]. The Affective Slider is a derivation of the SAM that uses a slider scale instead of the five point Likert scale [6]. The Affective Slider has been validated to deliver same results as the Self Assessment Manikin and it's advantages include the easy reproducibility in current digital devices [6].

To summarize, there is a clear difference between the quality of the raw data provided by the Emotiv Epoc+ headset and laboratory level EEG measurement tools. Still, the headset has been used in several studies, but not in for example medical context, in which data quality is highly important [2]. At the same time validated tools for measuring the subjective level of arousal exist, like the Affective Slider [6]. To test the quality of the live metrics provided by Emotiv, we formulate our research question in the following way:

Does the Engagement/Boredom level provided by the commercial software Emotiv Epoc+ SDK correlate with the arousal level assessed by the user?

Our null hypothesis is that the Engagement/Boredom level captured by Emotiv headset correlates with the arousal level evaluated with the Affective slider.

The Emotiv headset comes with a convenient Emotiv Control Panel software that processes the signals recorded by the Emotiv Epoc headset in real time. The quality of the live metrics has not yet been studied by third parties, hence we create an application to record the Emotiv live metrics and corresponding data using the Affective Slider. We show test participants images from dataset of affective images [7], after which they were asked to rate their pleasure and arousal level, while recording their EEG with the Emotiv headset. Finally we compare the data from the Affective Slider and Emotiv headset to find out if they correlate. We use Processing coupled with the Emotiv Xavier SDK to create the interface of our test as well as capturing the data for the analysis. The analysis will be done in the second phase of the project after running the tests.

2. RESEARCH METHOD

Our study consists of an experiment carried out in a laboratory setting with 14 participants. Each user was asked to rate their response to 60 images with the Affective Slider while measuring their excitement level with the Emotiv Epoc+. Afterwards the data was tested for a correlation using Spearman's correlation.

2.1 Experimental Setting

The experiment was conducted in a laboratory setting at TU Berlin. Each experiment was conducted in the same experiment room, so that participants could focus in silence without interruptions or distractions. During the experiment some pictures were taken with the consent of the participant.

The experiment focused on measuring simultaneously the participant's excitement level with Emotiv Epoc+ and the arousal level with the Affective Slider. For measuring real-time brain activity from the Emotiv Epoc+ the Java version of the Emotiv Xavier SDK version 1.0.0.3 Premium was used. All images shown to participants were taken from the International Affective Picture System randomly. The participant facing interface queries the users about their feelings for each picture with the Affective Slider, and finally saves the data from the Emotiv headset and Affective Slider for later analysis. This application was implemented with Processing.

2.2 Participants

As participants in this study students from the TU Berlin campus were recruited. We posted recruiting advertisements on the class forum and on several social network platforms, with basic information about the study and a note that no compensation was given. The people who were interested in participating in the study could book a slot of around 30 minutes with an online calendar platform. 14 users participated in the experiment, including 5 females and 9 males, all students, and an age range from 21 to 28 years old (Table 1. Participant Demographic). In an additional questionnaire [Appendix 2], participants were asked about their sleep the day before the test, as well working/studying hours per day, coffee and smoking habits in their daily life. This data was collected in

case problems arise in the data collection with the Emotic Epoc+.

Table 1. Participant Demographics

Age Range/years old	21-28	
Gender	Female	5
	Male	9
Background	Student	14
	Employee	0
Smoker	Yes	5
	No	9
Sleeping time/hours	4.5-9	

2.3 Measures

Two variables were measured during the test. The first one is the participant's excitement level measured with the Emotiv Epoc+, and the second one is the self assessed level of arousal of the participant using the Affective Slider. Both variables were in a form of a floating point value between zero and one. Each user was shown 60 randomly selected pictures out of set of 800 affective images [7], out of which the first 10 pictures were treated as training data for reducing any noise in the data. Participants were informed about the first 10 pictures being training. Other variables were not changed during the experiment.

2.4 Procedure

At the beginning of each test, the charge on the Emotiv Epoc+ was checked, and all sensors of the device were made wet before the participant arrived. To provide more friendly user experience, the two measurements were done on different laptops. Therefore, the clock time on both test laptops were synchronized to the same second, to remove the any influence of time difference.

In the first place, all participants were asked to fill out the consent form [Appendix 1]. After that they were explained the procedure of the experiment, including the meaning of the Affective Slider and how to scale the slider by using the touch screen on our experiment laptop (Figure 1). Additionally, we provided instructions of how to use the Affective Slider on paper next to the experiment screen. Once the participant was ready for the test, they could start by clicking the start button. Simultaneously the Emotiv Epoc+ measurement was launched. Participants were asked to focus on each picture at least 2-5 seconds, then scale the slider on the screen according to their feeling. Moving to the next picture was done by clicking the next button, which was always set in the same position of the screen. Further, there was a reminder shown on the next button indicating how many pictures the participant had already seen (Figure 2).

Figure 1. User test Process. Measurement from Emotiv Headset and Affective Slider at the same time

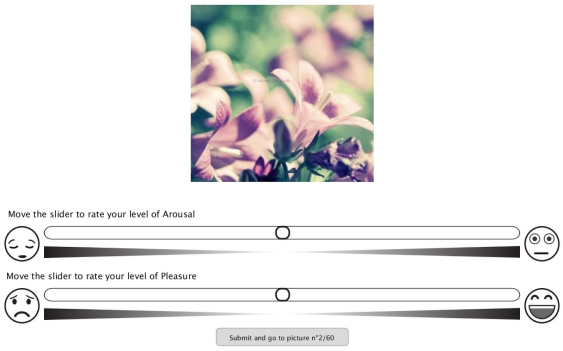

Figure 2. Testing Prototype Interface with Affective Slider

Once the participants finished rating their state for all 60 pictures, a finishing notification was shown on the screen in text. Then, the Emotiv headset was removed from the participant's head, and the real-time measurement on Emotiv Epoc+ was stopped. Finally, all the participants were asked to finish a questionnaire [Appendix 2], which was completely anonymized, like all the data gathered during the experiment.

2.5 Analysis

The analysis was conducted using Python. We

executed a linear correlation analysis between two variables: the Engagement/Boredom level provided by the software Emotiv SDK, and the level of Arousal assessed by the user with the Affective Slider.

Engagement/Boredom level from Emotiv software is provided 3 times per second. To ensure that the score represents the user genuine reaction to the image displayed, the data has to be processed. The method adopted is a 1 second value average after the display of the image. A synthesis of the course of events during the experiment can help to understand the data processing executed.

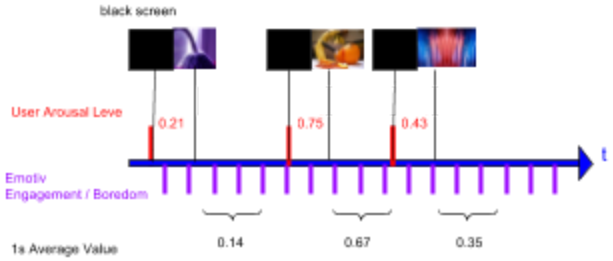

Figure 3. Synthesis of the course of events in the experiment

1. An image is displayed to the participant. The participant can observe it as long as he wants.
1. The participant assesses his Arousal Level.
2. A black screen is displayed during 3 seconds

Then, these variables have to be standardized, in order to allow direct comparison. After normalization, the absolute error between the scores is calculated. These are compiled in figure 4. Finally, a linear correlation is applied to the datasets, for each series of samples corresponding to a participant. Indeed, the test concerns the capacity of the Emotiv headset to assess the emotions of a user.

3. RESULTS

Concerning the quality of the data, the range of the absolute error between the two normalized variables is quite large (figure 4). There is some isolated points. This shows that the quality of the data obtained with Emotiv headset is not optimal. It may contain some unwanted artifacts.

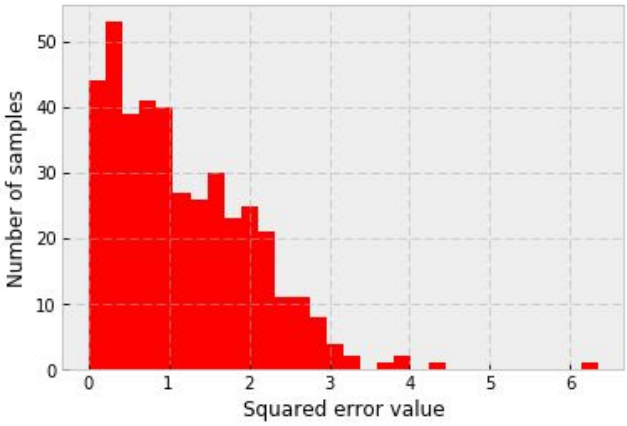

Figure 4. Error range distribution.

Average Correlation : -0.01
Average p Value : 0.43
Number of points: 449

The results from the linear correlation calculated using the data from all the 50 pictures for each user can be found in Table 2. For two participants a correlation could not be calculated due to a hardware error in the experiment resulting from a disconnection to the Emotiv Epoc+.

Table 2. Linear Correlation Results

Participant	Correlation	p value
1	0.25	0.08
2	0.03	0.86
3	0.23	0.16
4	0.0	0.99
5	-0.13	0.36
6	-0.16	0.33
7	-0.14	0.35
8	non exploitable	non exploitable
9	-0.09	0.65
10	non exploitable	non exploitable
11	-0.06	0.66
12	0.09	0.53
13	-0.14	0.34
14	-0.07	0.64

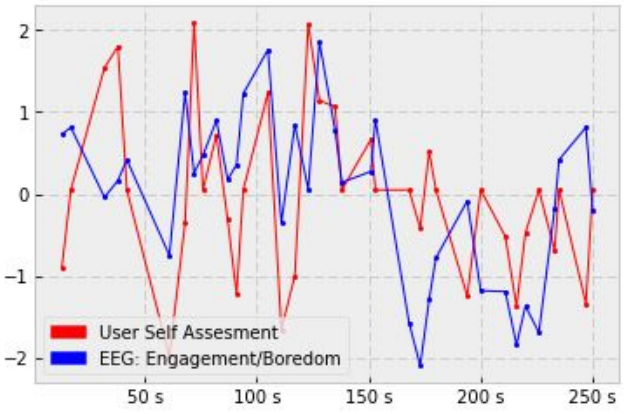

Figure 3. Result visualisation for one single user, on time experiment

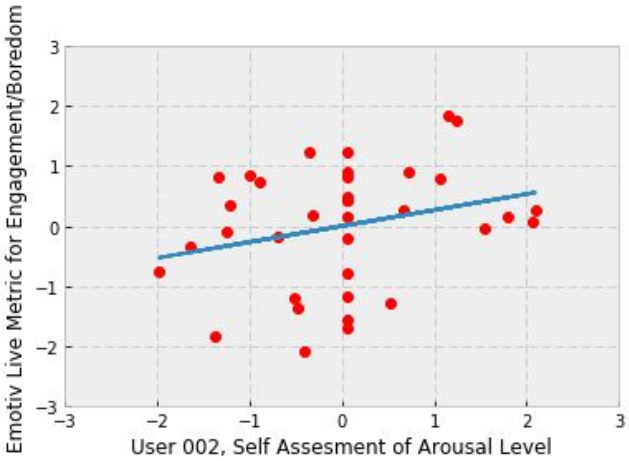

Figure 4. Result visualisation for one single user, and linear correlation
correlation = 0.27 / pvalue=0.12

A cleaning of the data has been executed, in order to see if the results may have been deteriored by a systematic error during the experiment. The method used has been to eliminate the data with an absolute error above 2. This resulted in the following results:

Average Correlation : 0.29
Average p Value : 0.21
Number of points kept after data cleaning : 362
Percentage of points after data cleaning : 57%

4. DISCUSSION

As shown in the results, in average no correlation was found between the Emotiv Epoc+ live metrics. So the answer to the research question, whether the Engagement/Boredom level provided by the commercial software Emotiv Epoc+ SDK correlates with the arousal level assessed by the user, is no, and hence our hypothesis was proven wrong. However, when looking at smaller chunks of data recorded during the study, better correlations than the average value were found. This is an indication that in average the measurements captured by the Emotiv headset are very noisy. This is also well inline with the previous studies [1] [3] [4], that show that the raw data captured by the Emotiv Epoc+ headset is noisy and not on the same level with laboratory grade devices.

All in all the results suggest that the Emotiv Epoc+ live metrics should not be used in critical application such as medical or research applications. However, the headset is still suitable for less critical applications, like gaming. During the experiment some limitations of the Emotiv Epoc+ headset were observed to have some effect on the results. Firstly, the headset doesn't fit every participant, as the some people have a head so small the individual sensor on the headset are not well connected with the scalp. Secondly, not every sensor of the headset was working well all the time as the noise level for each sensor on the headset varies a lot even when the user is sitting silently and not moving. Hence the data is at points very noisy and also influences the reliability of the live metrics. In addition, the data analysis part was limited by the polling rate of the Emotiv Epoc+ for the live metrics. The headset only allows to query the live metrics only about three times a second, which limits the temporal details of the results and the analysis of the results. As the temporal resolution is not high, there is always a small shift between seeing a new image and the time window out of which the average engagement/boredom level was calculated. This effected the synchronization of data from the Emotiv Epoc+ and the Affective Slider by reducing the accuracy of calculating the correlation.

The limitations of our study include the quite small sample size of 14. In addition, a one time measurement is not optimal, since the Emotiv SDK

requires a user profile to be created for each new user and uses some of the previously collected data to adjust the results. Due to compatibility reasons the study was not conducted with the latest version of the Emotiv SDK, which would have also provided the live metrics for the valence level of the user. This would have provided an additional data point, as the valence level was already measured with the Affective Slider. Also, using other stimuli, like music, could have created stronger reactions, so this is a possible future improvement of the experiment. Finally, the Affective Slider was used differently by the participants, as some used the whole scale, while other only used the middle range. A balancing factor for this effect would need to be added.

5. FUTURE WORK

The possible future work contains three directions. Firstly, as the Emotiv SDK used in the experiment is not the current latest version, the version would need to be updated to also get the valence level from the Emotiv Epoc+. Secondly, different stimulus could be used to trigger brain activity of participants, for instance, music or showing videos. Lastly, having more participants in the experiment could be beneficial to figure out if these changes could improve our experiment result or not.

To define the level of noise in the data, the noise level could be measured with the help of motion sensors, in order to focus on the effectiveness of the live metrics based on noiseless data. This does however not fully serve the current research motivation, as the aim was to validate the live metrics in an environment that it is marketed to be used in normal consumer environment.

6. CONCLUSION

New EEG sensors aimed at the consumer market are being launched. These sensors that are considerably cheaper than the laboratory grade research devices, faster to set up and provide live analysis of the EEG data, and hence provide an interesting possibilities for new application areas for using EEG. One of these sensors is the Emotiv Epoc+ that provides live metrics of the user's arousal level. The quality of the live metrics has so far not been validated, studies have only addressed the quality of the live data captured by the Emotiv Epoc+ [1] [3] [4]. As using the live metrics provide great convenience for applications that need the user's valence or arousal level, this study focused on validating the quality of the live metrics. The results show that in average the Emotiv Epoc+ live metrics do not correlate with the user's self assessment, but in smaller chunks of data some correlation can be seen.

This study only validated the arousal level provided by the Emotiv Epoc+ live metrics and hence further studies would be needed to validate the other measures provided by the live metrics. However, the results are well in line with the previous literature stating that the data provided by the Emotiv Epoc+ is noisy [1] [3] [4].

All in all the results shows that the live metrics from Emotiv Epoc+ sensor, although being convenient to use and quick to set up, should not be used in critical applications, as the the quality is not well correlated with the user's self assessment. The quality of the Emotiv Epoc+ live metrics should only be used in non-critical applications, like for example gaming.

REFERENCES

[1] Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., and Dutoit, T. 2003. Performance of the Emotiv Epoc headset for P300-based applications. *BioMedical Engineering OnLine* 2013

[2] Emotiv. 2017. EPOC+. [online] Available at: https://www.emotiv.com/epoc/ [Accessed 17 Feb. 2017]

[3] 3 Badcock, NA., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., and McArthur, G. 2013.Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 1:e38. DOI=https://doi.org/10.7717/peerj.38

[4] 4 Stytsenko, K., Evaldas, J., and Cosima, P. 2011. Evaluation of consumer EEG device Emotiv EPOC. *MEi: CogSci Conference 2011*, Ljubljana. 2011.

[5] 5 Bradley, M., Lang, P. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*. 25:1, 1994, pp. 49-59. DOI=http://dx.doi.org/10.1016/0005-7916(94)90063-9

[6] 6 Betella, A., and Verschure, P. 2016. The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLoS ONE* 11, 2. DOI=doi:10.1371/journal.pone.0148037

[7] 7 Machajdik, J. 2017. Affective Image Classification Using Features inspired by Psychology and Art Theory. [online] *Imageemotion.org*. Available at: http://www.imageemotion.org/ [Accessed 18 Feb. 2017].

[8] 8 SciPy.org. 2017. scipy.stats.spearmanr. [online] Available at: https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html [Accessed 18 Feb. 2017]

APPENDIX 1 EXPERIMENT CONSENT FORM (1 / 2)

Neuro-usability, Winter semester 16/17 TU Berlin, Quality and Usability
Participant No.

Emotiv Epoc+ Live Metrics Quality Validation Test

Introduction
We kindly ask you to partake in a scientific investigation, which is carried out by the course of Neuro Usability at TU Berlin. The group to contact about this research with any question or comment you might have is following, we are all master students at the TU Berlin. Feel free to contact us by email or tel:

Yuan Lyu
+49 157 3326 2956
lvyuan0302@hotmail.com

Mikko Honkanen
+49 1715339560
mikko.honkanen@campus.tu-berlin.de

Stéphane Garti
+4915733262135
stephane.garti@campus.tu-berlin.de

Purpose of the study
The purpose of this research is to record your brain activity during the with a testing application with a PC laptop by using Emotiv Epoc+ device.

Inclusion criteria
Inclusion criteria for participation in the study are consenting healthy brain, not blind.

Procedure and duration of participation:
In the study you need to scale your affective states on pleasure and arousal by each picture we provided, at the same time, we will record your brain activity by wearing the Emotive Epoc+ device. All the brain data will be recorded by the Emotive Control Panel software but only valence and excitement values will be used in our research. In total, you will watch 60 pictures and fill 60 pair of affective sliders, the first 10 pictures' data will not be used in the validation but as a training. After the test, you will be required to finish one simple questionnaire.

Potential risks of participation:
There are no particular risks associated with you partaking the study.

Circumstances leading to discontinue your participation
Should you at any point in time wish you change your mind about your willingness to partake and decide to withdraw from the study or to not follow the instructions of the researcher, the experiment and your participation will be discontinued immediately without consequences.

Confidentiality and data privacy protection:

APPENDIX 1 EXPERIMENT CONSENT FORM (2 / 2)

Neuro-usability, Winter semester 16/17 TU Berlin, Quality and Usability
Participant No.

By signing the consent form you declare your willingness to partake and agree for the group of researchers involved in the study to collect and process your data for the purpose of the study. All of your data will be anonymized, even the questionnaire you filled in the final of the test, after that these will be used for research and statistical analysis together with your brain data. In the event of scientific publication of the results of the study, you will by no means be identifiable and you data remain completely anonymized. Anonymity will be assured by assigning a subject number to all of your data. We might take some pictures during the testing but your face will not be recorded.

You have the right to be reported back all of the data collected from you by the researcher. You also have the right to correct any in accurateness in our data. In the event of any question or comment of yours, please contact the lead researcher, whose contact details are included in this participant information sheet. You have the right to object to the further processing of your data within the course of the study and request their deletion up to two weeks after your participation. Please contact the leading researcher in case of questions and state your subject number. You can find you subject number on the top of your copy of this document

Right of knowledge/ duty of notification:
You have the right to inquire about any aspects of the study at any point of time. You as the participant have the duty to follow the instructions given by the researcher for duration of the participation in the study and to inform the researcher in the event of any change in your health or well-being.

Voluntariness of participation
You have the right to discontinue participation in the study without informing the researcher about the reason; this includes the right to withdraw your willingness to participate and you right to withdraw your agreement to data collection and processing, without any consequence for yourself.

Signature:

Date:

APPENDIX 2 EXPERIMENT QUESTIONNAIRE

Neuro-usability, Winter semester 16/17 TU Berlin, Quality and Usability
Emotiv Epoc+ Live Metrics Quality Validation Test

Questionnaire

Participant No.____

1. Age:
2. Gender:
3. Are you student or employee?
   - ❑ student
   - ❑ employee
4. How many hours do you study/work per day?
   - ❑ 1-2
   - ❑ 2-5
   - ❑ 5-10
   - ❑ 10+
5. What time did you wake up today?
6. How many hours did you sleep last night?
7. Are you a smoker?
   - ❑ Y
   - ❑ N, skip question 8
8. How many cigeratte do you smoke per day?
   - ❑ 1-2
   - ❑ 2-5
   - ❑ 5-10
   - ❑ 10+
9. Do you usually drink coffee?
   - ❑ Y
   - ❑ N
10. How often do you have coffee per day?
    - ❑ <1times
    - ❑ 1-2 times
    - ❑ 2-5 times
    - ❑ 5 times+