ISASE 2019

A User Evaluation System Using Sensors of Smartphones

Ryo Yamaguchi*, Jue Zhang** and Naiwarla P. Chandrasiri***

** Kogakuin University, 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo, Japan*
em17022@ns.kogakuin.ac.jp
*** Kogakuin University, 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo, Japan*
zhangjue@cc.kogakuin.ac.jp
**** Kogakuin University, 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo, Japan*
chandrasiri@kogakuin.ac.jp

Abstract: The five-point scale is a method of expressing assessments that are fundamentally continuous in nature as five discrete points. Though quantitative data is a necessity for statistical analysis, it is possible that this does not entirely accurately reflect the feelings of those surveyed. We considered that in order to avoid that possibility, it may be possible to substitute fundamentally unquantifiable measurements, such as emotional assessment, with more quantifiable ones as a method of more intuitively and quantifiably measuring feelings. To that end, we wondered if it might be possible to use the sensors in people's smartphones as a ubiquitous and always close-at-hand device that is capable of making quantitative measurements. In this research, we focus on the sensors with which smartphones are equipped, proposing, and investigating the effectiveness of, the measurement and numerical analysis of the force with which a device is shaken as a more direct means of conducting emotional assessment and that succeeded.
Keywords: *kansei evaluation, smartphone, sensor*

1. INTRODUCTION

The five-point scale is a method of expressing assessments that are fundamentally continuous in nature as five discrete points. It is widely used as a means of making assessments discrete for the purpose of easy statistical analysis, etc. The five-point scale can be used to quantitatively and qualitatively analyze and draw data from complex, unclear feelings, and to perform comparative analysis between several test subjects/samples. Though the discrete values provided by the five-point scale allow for easy statistical analysis, some state that with only five points through which to express their feelings, they find it very difficult to decide on a value. Though quantitative data is a necessity for statistical analysis, it is possible that this does not entirely accurately reflect the feelings of those surveyed. We considered that in order to avoid that possibility, it may be possible to substitute fundamentally unquantifiable measurements, such as emotional assessment, with more quantifiable ones as a method of more intuitively and quantifiably measuring feelings. To that end, we wondered if it might be possible to use the sensors in people's smartphones as a ubiquitous and always close-at-hand device that is capable of making quantitative measurements. In this research, we focus on the sensors with which smartphones are equipped, proposing, and investigating the effectiveness of, the measurement and numerical analysis of the force with which a device is shaken as a more direct means of conducting emotional assessment.

2. EXPERIMENT INTO ACQUIRING SHAKING DATA

2.1 Experiment overview

There are two main types of motion-detecting sensors with which devices are equipped; accelerometers and gyro sensors. In order to answer questions such as whether we should take the maximum, average, or frequency value of these devices, whether there would be a significant difference in strength of shaking between different individuals, etc., we conducted a preliminary experiment investigating 1. Whether we should use the accelerometer or gyro sensor, 2. Which value we should take of the maximum, average, and frequency, 3. Whether there was a difference in the strength of shaking between individuals.

We developed a measurement application that used the sensors with which Android devices are equipped to conduct our experiment. Our sample consisted of 20 students in their 20s (14 male, 6 female). We established three conditions and had participants shake their devices for us.

2.2 Experiment methodology

The three conditions were whatever the participants personally felt to be "Weak," "Medium" and "Strong" shaking. Participants shook their devices ten times for each condition, for a duration of five seconds each time.
We took values from the accelerometer and gyro sensors, which are both capable of reading the device's movement. We took both the maximum and average values from both sensors, as well as the frequency of the accelerometer, for a total of five values.

2.3 The measuring application algorithm

We drew three-dimensional data from the accelerometer and gyro sensor. The sensors were sampled approximately 50 times per second.

The average was calculated by summing all the values sampled from beginning to end of the sampling period, then dividing that number by the number of samples taken. For the maximum value, a variable, "max," was initialized at 0, and set to the last value sampled if its current value was lower than that value. The frequency was calculated by initializing a variable, "count," at 0, then, if the sign of any of the accelerometer's three-dimensional values (x, y, z) had changed from the previously recorded value, incrementing "count." "count" was then divided by 6, such that its value represented the number of completed cycles on all 3 axes.

2.4 Experiment results and considerations

We conducted correlation analysis on the normalized data drawn from the five measuring methods used in our experiment and the categories of Weak, Medium and Strong, expressed as -1, 0, and 1 respectively. We found that the highest correlation coefficient existent between the force of the shaking and the average acceleration (Chart 1).

Chart 1: Correlation coefficient and results of significance test (acquiring shaking data experiment)

Correlation coefficient				
Maximum Acceleration	Average Acceleration	Maximum Gyrovalue	Average Gyrovalue	Frequency
0.848	0.881	0.779	0.782	0.345
				Max value
significance test				
Maximum Acceleration	Average Acceleration	Maximum Gyrovalue	Average Gyrovalue	Frequency
0.000	0.000	0.000	0.000	0.000
		1:0.001	Yes	No

Since average acceleration was the measurement method with the highest correlation, we calculated the standard deviations for Weak, Medium and Strong shaking against the average value of average acceleration per-person, the average acceleration values for the entire sample group, and those values according to gender, as shown in Chart 2. Looking at the whole-group values, we see that as the as the force of shaking increases, so does the average and SD of each participant's average acceleration. When looking at results by gender, we see that the men's values exceed the individual average, and the women's values, while lower than the individual average for Weak shaking, exceed it for Medium and Strong, with Strong being a particularly striking difference. We can therefore conclude that, regardless of gender, as the force put into shaking increases, we reach a higher number.
From these findings, we concluded that we should set up a learning function that could account for individual differences by establishing defined levels before conducting assessment.

Chart 2: standard differentiation of Average acceleration

Standard Differentiation				
	participant's average	Whole-Group	men	women
Weak	1.369	2.525	2.681	0.940
Medium	1.369	3.585	3.748	1.948
Strong	1.499	4.972	4.907	4.741

3. APPLICATION OVERVIEW

We built our application using average acceleration as our index. First, we carried out user registration in which the application learned the strength of the user's force (i.e. the strength of their average acceleration) by having each user perform five sets of shaking, where a single set involved shaking at Weak, Medium and Strong levels of force. The initial two sets were not actually recorded. Pushing "Start" on the main screen caused the application to begin measuring. The average acceleration data from the shaking periods (hereafter called the "current data") was taken and temporarily recorded. The average of the "Weak average" and "Medium average" was taken as Level 1, while the average of the "Medium average" and "Strong average" was taken as Level 2. The value of the current data was translated to assessment of 0-25% when its value was between 0 and Level 1, 25-75% when its value was between Level 1 and Level 2, and 75-100% when its value was between Level 2 and the maximum value. The values of the current data were stored in the

Keywords: kansei evaluation, smartphone, continuous quantity

database as shown in Fig. 1 (the blue, green and red sections corresponding respectively to Weak, Medium and Strong), deepening our base of knowledge.

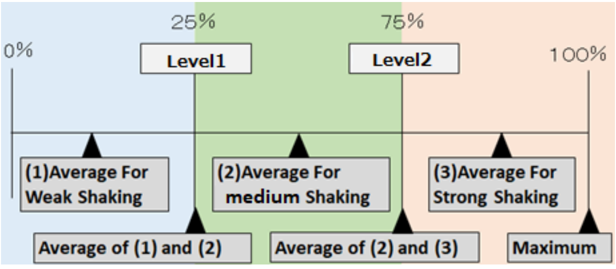

Fig. 1: The assessment application's algorithm

4. ASSESSMENT EXPERIMENT USING THE ASSESSMENT APPLICATION

4.1 Experiment overview

Based on the results of our preliminary experiment, we used the averages of the values drawn from the accelerometer to develop the assessment application as indicated in section 3, then conducted an experiment to check if the assessments provided by the application were in agreement with individual feelings.

Our sample consisted of 15 students in their 20s (10 men, 5 women). On this sample, we conducted an experiment consisting of the following two points.

1. The participant was asked 15 questions and asked to answer each of them on paper on a five-point scale.
2. The participant was asked the same 15 questions, but asked to answer them using our application.

We then conducted a brief interview regarding any points of interest.

4.2 Experiment methodology

Since carrying out two assessments in succession may have caused the first to influence the second, participants were numbered, with the odd-numbered participants carrying out the five-point assessment followed by using the application, and the even-numbered participants using the application followed by the five-point assessment.

The questions consisted of two types; participants were shown images of either 1. common foods or 2. landscapes, and asked how much they liked them.

4.3 Experiment results and considerations

In order to ascertain whether there was a correlation between the assessments drawn from the five-point scale assessments and the assessments using the application, we conducted correlation analysis of each group of assessments (whole group, odd-numbered participants and even-numbered participants). Each group demonstrated a high level of correlation (Chart 3). Since the five-point scale and our application demonstrated similar assessment results, we can conclude that our application's algorithm was effective at conducting emotional assessments.

Further, those who conducted the evaluation using the application first showed a higher level of correlation than those who used the five-point test first. From this we conclude that assessment using the application asserted a greater level of influence over the subsequent test than did assessment using the five-point scale.

Chart 3: Correlation coefficient and results of significance test (assessment experiment)

Correlation coefficient			significance test		
Whole-group	Odd number	Even number	Whole-group	Odd number	Even number
0.829	0.795	0.859	0.000	0.000	0.000
		Maximum	t<0.001	Yes	No

Chart 4: Average values and standard differentiation for each five-point scale assessment

Five-point scale	1	2	3	4	5
Average	15.729	28.082	45.02	69.179	87.990
Standard Differentiation	4.513	14.485	20.706	18.190	12.861

Calculating the average values and standard differentiations for each application assessment against the corresponding five-point scale assessment, we found that the majority of standard differentiations were close to 3 on the five-point scale (Chart 4).

From these results, we conclude that the five-point scale is less likely to represent middling assessments than is our application.

Further, from interviews, we received views such as "Doing the assessment by shaking my phone was more fun than the five-point scale," and "I get more and more tired from shaking my phone."

5. EXPERIMENT TO IMPROVE THE APPLICATION

5.1 Experiment overview

In Section 2, we decided on the accelerometer after concluding that it better represented the force with which a device was shook than did the gyro sensor. However, it was clear that the gyro sensor's values *did* better reflect shaking force for a minority of participants. Our next experiment, therefore, aimed to clarify the role of the gyro sensor by finely measuring the strength of shaking and thus more finely distinguishing its traits. At the same time, we also investigated methods of reducing the fatigue caused, and whether the order of evaluation made any difference. Our sample consisted of 14 students in their 20s.

5.2 Experiment methodology

We conducted our experiment using a slightly modified version of the application described in Section 3.

Participants shook their devices according to two conditions:

1. Three categories of force: Weak, Medium, and Strong (15 times)
2. To indicate numbers 1 through 10, where 1 was the weakest and 10 the strongest. (20 times)

For condition 1, participants shook their devices for five sets, where one set consisted of shaking their phone at strengths of Weak, Medium and Strong in that order. In addition to allowing participants to get used to the act of shaking their phones, this condition was conducted in order to investigate whether it was possible to reduce the number of sets used by the application in Section 3 in order to reduce fatigue.

For condition 2, odd-numbered participants shook their phones at a level of force corresponding to the numbers 1 through 10, then in reverse order, from 10 through 1. Even-numbered participants shook their phones at a level of force corresponding to the numbers 10 through 1 in reverse order, then from 1 through 10. In prior experiments, we had used a three-point system of Weak, Medium and Strong at this stage, but we established a ten-point system in order to gather more detailed data. The process was conducted in two different orders in order to investigate whether the previous force of shaking had an impact on the force of subsequent attempts.

5.3 Our application

The main differences from the application described in Section 3 were:

1. Strengths 1 to 10 were assigned to the accelerometer and gyro sensor data.
2. The value that had been displayed when measuring was now hidden.

Change 1 was made in order to measure the strength of shaking as a ten-point scale. Change 2 was made so that the strength of participant's shaking would not be influenced by the displayed value.

5.4 Experiment results and considerations

We calculated the correlation coefficients of the strengths 1 through 10 and the data collected from the four categories of average acceleration, maximum acceleration, average angular velocity and maximum angular velocity according to the conditions described in 5.2. We found that for the majority of participants, average acceleration showed the highest correlation, while still showing a high level of correlation among the remaining participants. We therefore concluded that average acceleration still best represents the strength of shaking when using a ten-point scale. On average, average angular velocity showed the 2^{nd} highest correlation; no less than 0.8 regardless of participant. We believe this value could be used in place of average acceleration.

In order to investigate what the result would be if we were to take only one, rather than five sets of learning data in order to reduce fatigue, we separately took the first and second sets of data collected according to Condition 1 described in 5.2 and used them individually as learning data. Then, using the algorithm described in Section 3, we assessed the 20 points of data gathered as described under Condition 2 of 5.2, and calculated the correlation coefficients of those assessments against the actual strength values. We found that both Set 1 and Set 2, showed a high level of correlation regardless of whether the participant was even- or odd-numbered. We therefore conclude that it would be possible to use just a single set of learning data. When conducting a one-tailed t-test at the 5% level comparing sets 1 and 2 to the even- and odd-numbered groups, we found no significance. We can therefore conclude that there is no major difference between the results of Set 1 and Set 2, and that Set 1 alone can therefore be used as learning data.

6. RESULTS AND FUTURE ISSUES

In this research, we succeeded in using smartphone sensors to create an assessment system with the potential to assess more intuitively, while not diverging significantly from five-point scale assessment, but allowing for more detailed assessment.

In future, it will be necessary to further consider methods of analyzing factors not fully analyzed in this research, such as the relationship between shaking force and each sensor, and the personal traits of each individual. We believe it would then be possible to devise an algorithm that is more accurate and better takes into account individual differences.