A study toward development of an assessment method for measuring computational intelligence of smart device interfaces

Ahmet Al Zarqa, Tarik Ozkul, and A. R. Al-Ali

***Abstract*—**Today smart machine market has reached enormous proportions. There is a strong drive among manufacturers to make software interface of “smart devices” as intelligent as possible to attract customers. So far the only feasible way for measuring intelligence of software interface has been conducting a survey among users to see how users rate the quality of the human-machine interface software. This study aims to develop an objective yardstick to measure intelligence of user interface based on observation of user behavior coupled with amount of interchanges between the user and the machine. The experiments indicated that user satisfaction of software interface predicted using this technique is very much in line with the paper based survey results reported by the actual users. This method may provide an objective way of measuring quality of user interface even before the actual users use the software.

***Keywords*—**User interface, Machine Intelligence Quotient, MIQ, User Intelligence Quotient, Computational intelligence

I. Introduction

TODAY mobile gadgets became an important part of our lives. The intent of this study is to find an objective way of measuring quality of user interface of smart devices. As smart phones and similar swipe activated tablet devices became part of our everyday life, the issue has started to gain more and more importance. Today’s demanding customers expect their mobile devices to act as “smart” as possible when they are trying to accomplish a task with them. It is expected that modern handheld devices assist their operators to accomplish complex tasks as easily as possible. This is essentially done by the user interface and the operating system of the mobile device. There have been several studies in the literature attempting to measure user interface quality of mobile devices [1]-[4].

In this study, a new approach to measure user interface quality is suggested. The starting point of this new approach is measuring “smartness” of the user interface. Indeed what users’ desire from the mobile gadgets is “smartness”. In order to explain the process, the term “smartness” need to be defined.

Fortunately there is some ground work done by researchers in defining “intelligence” of robotic machines. The terminology is initially developed for defining intelligence of robots, but the concept of “intelligence” can be easily extended to defining “smartness” of smart devices.

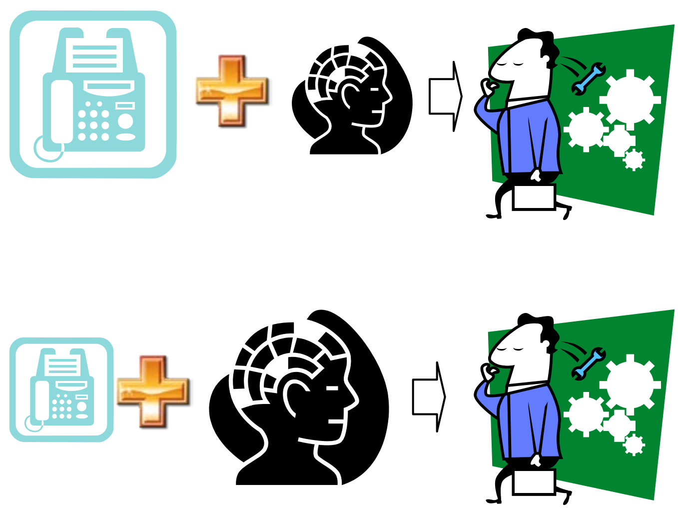

Fig. 1 Two different ways of completing a given task is shown. Top case shows an intelligent machine where operator needs to spend relatively less effort to do a task. Bottom case shows a less intelligent machine where operator needs to spend more effort to do the same task. Second machine has lower MIQ.

Machine Intelligence Quotient, or widely known as MIQ, introduces a philosophy for measuring intelligence level of machines. In this approach, “intelligence of the machine” is proportional to the level of help it provides for its master –human user- to complete his/her task. In this philosophy, tasks are for the human beings to tackle. Only human beings can initiate and finish a task. Machines can only help human beings to make their tasks easier. Intelligence level of the machines is graded according to the degree of help they provide to their human operator. Imagine two different scenarios where in the first case machine number one is very

A. Al-Zarqa is a graduate student in the Computer Science and Engineering Department of American University of Sharjah. PO Box 26666 Sharjah UAE. (e-mail: b00022890@aus.edu)

T. Ozkul is with the Computer Science and Engineering Department of American University of Sharjah. PO Box 26666 Sharjah UAE.(e-mail: tozkul@aus.edu, phone: +971 6 515 2455, Fax: +97165152979)

A. R. Al-Ali is with the Computer Science and Engineering Department of American University of Sharjah. PO Box 26666 Sharjah UAE. (aali@aus.edu)

helpful in completing a task and in the second case machine number two is only marginally helpful to its human operator to complete the task. As shown in Fig. 1, machine number one is considered to have more "intelligence" than machine number two. MIQ methodology defines an elegant way of determining boundaries of what human master do, and what machine does to help. End result of this process is an objective way of determining the amount of work done by the human master and the machine counterpart. More information about this process is available in the following references [5]-[8].

II. Overview of Technique Used

In this study, we have adapted MIQ philosophy to define the relationship between the human operator and the smart machine. In a typical case, human operator tries to achieve a task like finding a person's phone number from the contact number list to make a phone call or send an SMS message. Achieving this task requires human operator to go through different menu choices, make selections and then make further selections depending on the previous choices made. The human operator needs to initiate the task by tapping into appropriate buttons on the screen, and then understand the menu choices presented by the user interface on the screen to make further selections and entries. After several such iterations, user interface displays the phone number of the desired person on the screen and upon approval the call is placed. A simple task like this requires several rounds of interactions between the human operator and the mobile device. If the human operator achieves his or her purpose in the shortest possible time, with as few interactions as possible, this is attributed to the "smartness" of the mobile. If the same task is achieved with more steps and more iteration on a different mobile device, this is attributed to the smartphone being "less smart". This is in general how human operators rate the smartness level of mobile devices.

If the interaction process between the human operator and the smart phone examined carefully, the following factors appear to be effective in giving perception of "smartness" for mobile devices [9];

1. The number of steps required to achieve a given task, (e.g. How many steps required to achieve the task)
2. Ease of entering information into the smart device (e.g. Physical difficulty of swiping, entering info into device, size of buttons, sensitivity of the screen play a role in this.)
3. Ease of understandability of menu choices presented on the smart phone screen, (e.g. how cluttered the screen, how easy to decipher things on the screen, how understandable the icons are.)
4. Speed of response of the smart device. (e.g. after making an entry, how fast the device responds and displays the response screen.)

Today's discerning user is very demanding from the hardware as well as the user interface of the smart device. As far as users are concerned both hardware and software are considered contributing to the "smartness "of the smart device. Opinion polls of the users indicated that slow response rate of user interface, although it may be purely due to hardware issues, is considered debilitating to the "smartness" of the smart device.

The developed model takes into account all the factors stated above and generates an index which is proportional to the level of help provided by the user interface to the human user for completing a task. This unit-less index is called UIQ, "User Interface Quotient". UIQ is generated in an objective manner purely by observing and timing the response of the human user. Detailed explanation of the model will be done in the next section.

A. Formal Procedure for Calculating UIQ

The basic philosophy of determining the intelligence level of a smart system is given as follows [9]. A smart system is there to help a human operator to accomplish a task. Accomplishment of the task is the sole responsibility of the human operator and machines are there to help the human operator to accomplish the task. Accomplishments of a task require effort and intelligence from the operator. This effort is called "CIQ" which stands for Complete Intelligence Quotient for accomplishing a certain task. In case of smart phones, this may be an operation of sending an SMS to a person or entering a website using the smart phone. The human operator has to spend a certain amount of effort for accomplishing this task. The amount of human effort required toward accomplishment of this task is called "HIQ" which stands for Human Intelligence Quotient. If the human operator were to complete this task on his own with a simple bare-bone phone, this would be indicated as:

$$CIQ = HIQ$$

Now, if we consider the existence of a smart phone to help the human operator, then the equation would be modified as follows:

$$CIQ = HIQ + UIQ$$; which can be written as:

$$UIQ = CIQ - HIQ$$

In this equation, UIQ is the intelligence quotient of the smart phone helping the human operator. The smart phone helps the human operator to accomplish the task. As the smart phone is helpful in doing this, we consider that the work done the human operator becomes less (HIQ decreases and UIQ increases). When comparing the intelligence level of two different smart phones, we look into the level of help they provide to the human operator. The one that helps the human operator more in fulfilling the task at hand is considered to be the one with the higher UIQ.

In summary, the amount of total intelligence required for accomplishment of a specific task is constant. The smart phone that helps its operator to accomplish this task with the least amount of effort is considered to have UIQ (smart). The UIQ value is a relative index which indicates the level of smartness of two or more smart phones; the higher the value, the higher the intelligence.

As in the case of sending an SMS to a person, if the task can be done easier, with less effort on a specific mobile phone, the corresponding operating system (user interface) is considered to have higher UIQ.

B. Intelligence Task Graph

Any task that needs to be tackled can be represented by a series of interactions between the human user and the smartphone. Tasks can be represented by a task graph which displays various subtasks required to complete the whole task. The state diagram, as seen in Figure 2, consists of circles that represent the task and its complexity, and arrows indicated the flow of tasks from one to another. The diagram also distinguishes and shows the tasks that are completed by the human controller and those that are completed by the intelligent machine.

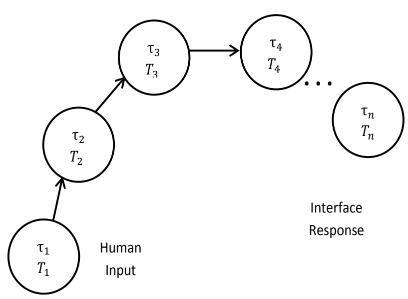

Fig. 2 Intelligence Task Graph

An intelligent task is made up of various smaller tasks, called subtasks, which are represented in a set called T:

$$T = \{T_1, \quad T_2, \quad T_3 \ldots T_n\} \tag{1}$$

Each subtask may have a different level of complexity, represented by τ :

$$\tau = \{\tau_1, \quad \tau_2, \quad \tau_3 \ldots \tau_n\} \tag{2}$$

Equations (1) and (2) show the set of subtasks and their complexities. As users try to tackle a subtask, there will be information transferred from user to machine or from machine to user. Data Transfer Matrix, which is designated as F, represents the amount of data transferred from one subtask to all the other subtasks during the completion of the main task. During the execution of the tasks, some tasks are handled by the smart device and some others are handled by the human operator. Task Allocation Matrix A indicates which tasks are handled by the human operator and which tasks are handled by the smart device. A and F are two matrices shown by equations (3) and (4):

$$F = \begin{bmatrix} 0 & f_{12} & f_{13} & f_{14} & \cdots & f_{1j} \\ f_{21} & 0 & f_{22} & f_{23} & \cdots & f_{2j} \\ f_{31} & f_{32} & 0 & f_{33} & \cdots & f_{3j} \\ f_{41} & f_{42} & f_{43} & 0 & \cdots & f_{4j} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{i1} & f_{i2} & f_{i3} & f_{i4} & \cdots & 0 \end{bmatrix} \tag{3}$$

In matrix F, fij represents the amount of data being transferred from subtask Ti to Tj.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} \end{bmatrix} \tag{4}$$

Matrix A has three columns. Column 1 represents tasks completed by the intelligent machine, column 2 represents tasks completed by the human operator, and column 3 represents the tasks completed by the non-intelligent machine component if there are any. Non-intelligent machine component is usually the basic electronics that take care of the phone functionalities like placing a call through GSM network.

C. Control Intelligence Quotient and Human Intelligence Quotient

Using the variables defined in the above section, the formula for "Complete Intelligence Quotient" (CIQ) and the "Human Intelligence Quotient" (HIQ) are given as follows:

$$\mathrm{CIQ} = \sum_{i=1}^{n} a_{i1} \cdot \tau_i + \sum_{i=1}^{n} a_{i2} \cdot \tau_i$$

$$\mathrm{HIQ} = \sum_{i=1}^{n} a_{i2} \cdot \tau_i + Cmh \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i1} \cdot a_{j2} \cdot f_{ij} + Chm_i = 1ni = 1n\, a_{i2} \cdot a_{j1} \cdot f_{ij}$$

In this formula, Cmh and Chm parameters define the "interface complexity values" that describes the difficulty of transferring data from a smart device to human and from human to smart device, respectively. The values of these variables indicate the difficulty of entering or interpreting data, and typically vary from 0 to 1. In this context "0" means little or no physical difficulty entering data into the mobile device, whereas "1" means extreme difficulty in entering the data to the mobile phone. To give an example, a mobile with extremely small keypad would have higher Chm than a mobile with large keypad. For well-designed system this parameter is expected to be around 0.05.

D. User Interface Intelligence Quotient (UIQ)

UIQ user intelligence quotient is the contribution of the intelligent device towards completion of a certain task. CIQ, complete intelligence quotient, is the total effort exerted by both human and machine together in completing the task. HIQ, human intelligence quotient, is the effort exerted by the human operator alone for completing the task. According to these definitions,

CIQ = UIQ + HIQ which leads to;

UIQ=CIQ-HIQ

UIQ is the work done by the user interface alone toward accomplishment of the task. Task graph is color coded to distinguish human and machine tasks. Gray colored tasks in Fig. 3 are performed by the smart device.

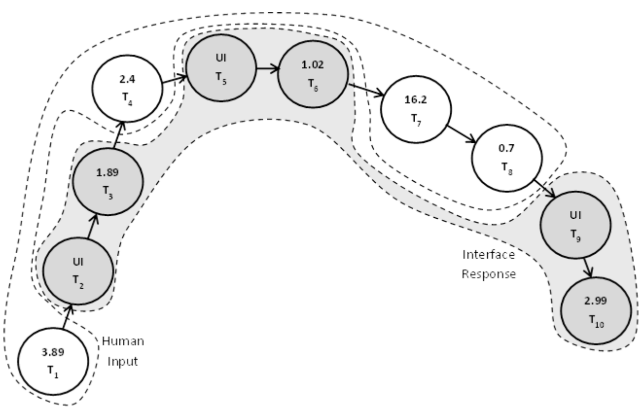

Fig. 3 Typical task graph is shown for accomplishment of a specific task. Gray colored circles indicate tasks done by the smart device.

III. Procedure for Experiments

The goal of the study is to find an objective procedure that generates a "smartness" score for the user interface. This score will be generated in an objective manner based on observations of the user. This score is expected to reflect the actual opinion of the user.

Experiments are conducted with 10 human operators who have similar age, education and dexterity level. Users are asked to accomplish several different tasks on a set of given smart devices.

The process is recorded so that important parameters are video recorded and parameters are extracted after the experiment is finished. Users are asked to complete the given tasks on different smart devices and asked to rate the relative "smartness" of the devices after the completion of the experiment.

UIQ index is calculated for each one of the tasks, for each one of the user using the procedure given in the previous section.

UIQ figures are compared to the user evaluated "smartness" figures in two different ways.

1. The raw index data generated by the UIQ procedure for the task is compared to user perceptions,
2. A fuzzy logic system is developed which uses the raw data of UIQ and generates an index of "smartness" based on a fuzzy model of smartness. Later this is compared to user perceptions.

IV. Fuzzy Logic System

Fuzzy logic system is designed to generate a number indicating "smartness" based on raw UIQ data. The system expects the following inputs:

1. Complexity of each subtask.
2. UIQ data
3. Total number of subtasks.
4. Cmh and Chm.

The complexity of each task is evaluated by examining the time it takes to complete.

These complexities account for the human input, as for the interface input it is handled by the interface response.

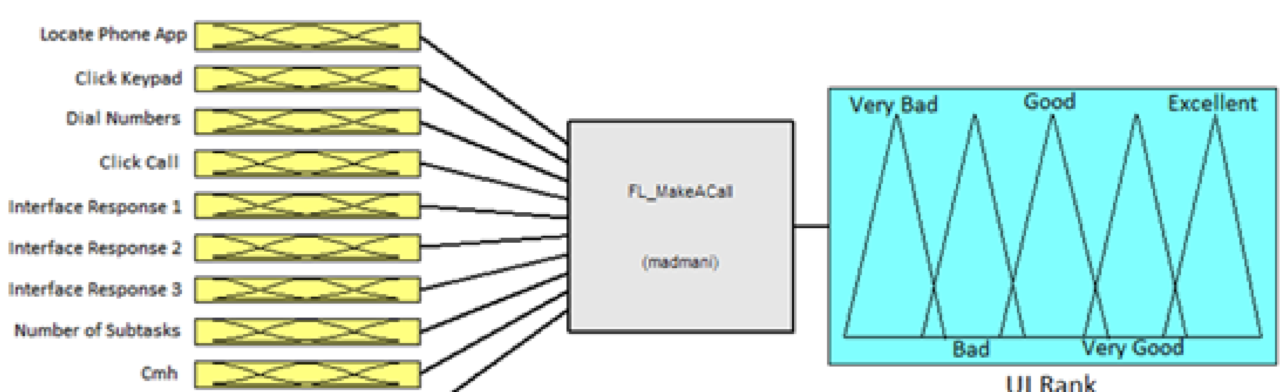

Fig. 4 FIS model for "Make a Call" Task

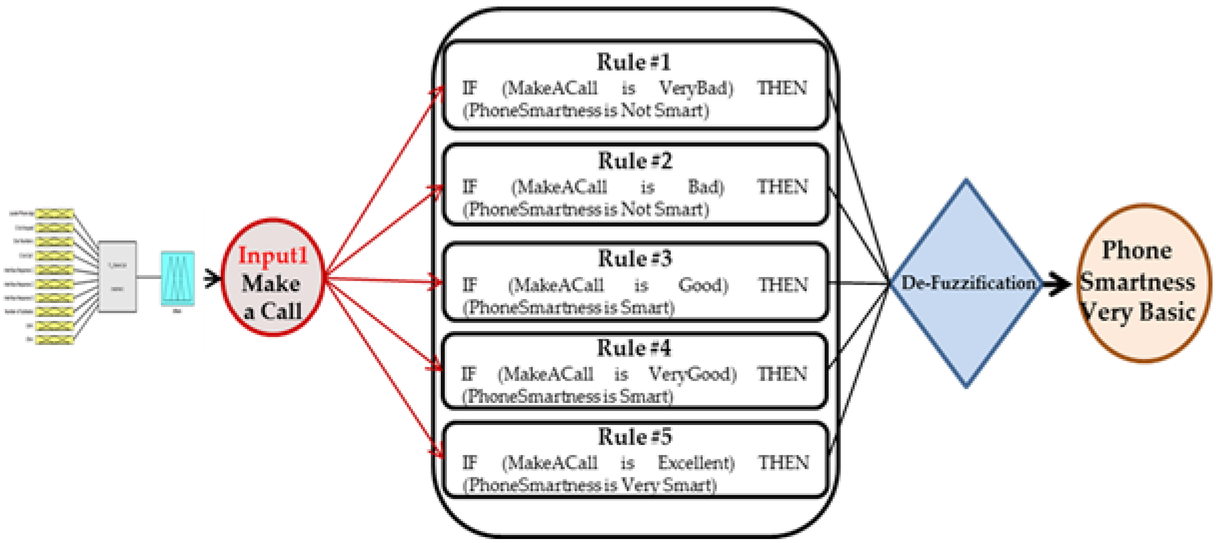

Fig. 5 FIS rules for interpreting quality of "Make a Call" Task

Each task can have a different number of subtasks with different smartphones. FIS system is used to generate one single output for a specific mobile device. Cmh and Chm values are also need to be entered into the FIS system, since they play a role in determining the smartness of the user interface.

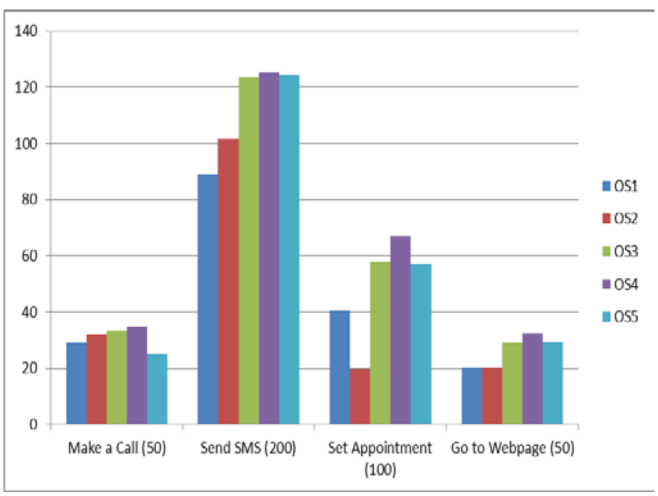

Fig. 6 UIQ results for 5 different smartphones, 4 different tasks.

Figure 4 shows the overall look of the FIS system. Figure 5 shows the some of the FIS rules. The output of the FIS indicates the "smartness rank" of the user interface.

V. Results of Experiments

User experiments are conducted with 10 different users, with five different smart phones (each with different operating systems) for four different tasks. Altogether 200+ different experiments were conducted to get the raw data reflecting interaction between smartphones and user reactions.

Due to excessive amount of data, only refined results will be provided in this section.

All of the smartphones are selected from commercially available units at the time of study. Each smartphone is selected to be working with a different operating system. Phones are selected from the most popular units available at the time of experiment. In order to avoid any commercial repercussions, names of smart phones and operating systems will not be revealed in the results. Since smartphones are associated with specific operating systems, only operating systems will be indicated in the results.

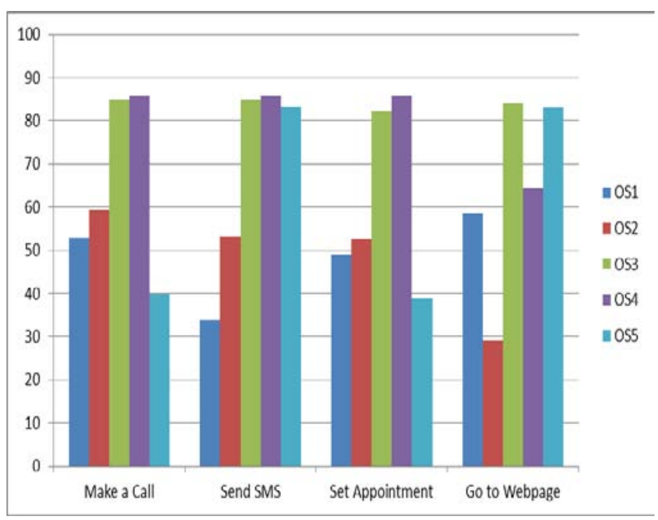

Fig 7 Results for Fuzzy Logic model

Figure 6 shows relative UIQ results of 5 different mobiles for 4 different tasks. The values are averaged for 10 users. The parameters were extracted from video recorded experiments while users interacted with the mobile every step of the way while performing the designated tasks.

Extracted parameters were inputted to the matrix formulas and presented below.

Figure 7 shows the results of the Fuzzy logic model. The raw data gathered from the video recorded experiments were inputted into the fuzzy logic model and generated results are displayed in the figure.

Figure 8 shows the survey results provided by the human users after conducting the experiments. The human users were asked to rank the "smartness", in other words "ease of use" of the mobile phone for performing the designated task. Each application is ranked by the human users individually. The results were averaged for the 10 users and displayed in Figure 8. The results indicated in this figure are the actual perceptions of the users after using five different mobiles for performing the tasks.

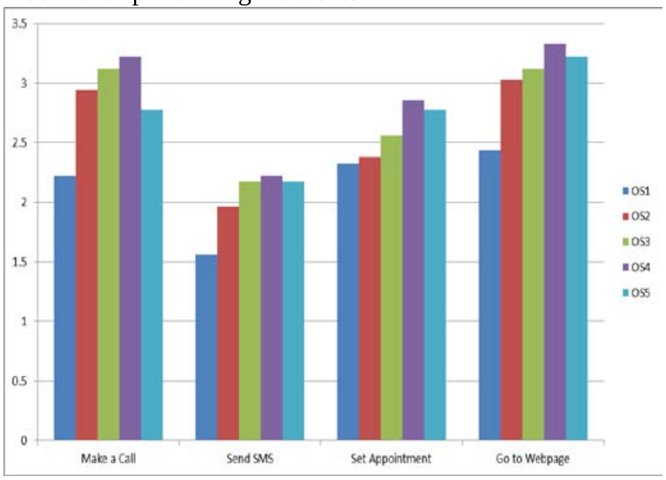

Fig. 8 Survey results reflecting opinion of the users about "smartness" of the smartphones.

VI. Conclusion

The results shown in Figures 6, 7 and 8 are refined results of 200+ experiments conducted with ten different users, using five different smart phones and performing four different tasks with each one of these smart phones.

Figure 8 indicates the opinion of the users regarding "smartness "of the mobiles for performing a given task. These results are found through user surveys and reflect how users feel toward the specific operating system while performing different tasks. Grading is relative and indicates that users found to perform some tasks easier in some of the smart phones. The results are averaged for all test users.

Figure 6 shows the results of UIQ methodology after extracting raw data from the experiments. The results are generated using observed and measured data from the experiments and calculated using the described UIQ process. The averaged UIQ results are found to be in parallel with the survey results shown in Figure 8.

Figure 7 shows the output of the fuzzy logic model which operates on the raw UIQ data. Rules of the fuzzy model are designed to reflect the user opinion. The results are generated by the FIS system based on data extracted through UIQ experiments.

The goal of this thesis has been to introduce a metric that can calculate the "usefulness" of a human-computer interface. The "usefulness" of the human-computer interface is commonly called "smartness" and becomes the center of attention with the proliferation of smart phones and smart machines. In today's aggressive smartphone market, user friendliness of the device, has been rated as one of the most important factor that wins customers. So far, the user satisfaction about the "usefulness' of the human-machine interface has only been determined by user surveys. Although, user surveys is the ultimate way of understanding the quality of the human-machine interface, an objective method of measurement, that is a metric, would be very useful for determining and measuring performance of our systems. From this point of view, the thesis may prove to be very useful for smartphone manufacturers, who would be able to run a series of tests using this metric and finally get numerical results that indicate performance of their system. In case the results are not as good as expected, they can come up with a different design before releasing the product to the public, hence saving a failure before it launches.

To the best of our knowledge there are no other objective methods that generate measurable metrics for "smartness" of the human-machine interface. In order to make sense out of generated metric, the "smartness" metric is presented in three forms which are complementary to each other.

a) Basic UIQ measurements which generate raw metrics for smartness of processes,
b) Fuzzy logic system output which generates a derived metric based on the smartness of multiple processes,
c) User surveys which indicate opinion of users in classical manner.

User survey results are used for determining fidelity of the results generated by basic user interface intelligence quotient (UIQ) and Fuzzy logic outputs.

Fuzzy logic was chosen as an alternative method of generating a metric since verbal description of quality of user interface has always been in words such as "good" and "bad," which is what fuzzy logic mainly deals with. Moreover, Fuzzy Logic made it easy to take the experiment another step further by calculating user interface intelligence quotient only for specific type of use.

The results showed that the UIQ metric and the Fuzzy Logic metric generated results that matched 70% of the times. The survey results indicated that UIQ and Fuzzy logic results have been within the ballpark, indicating being relationally correct. However, difference in actual values indicated that there is room for fine tuning of the membership functions of the fuzzy logic system.

As an explanation of the above process, the results for one of the tasks, Send SMS task, is shown in Figure 9.

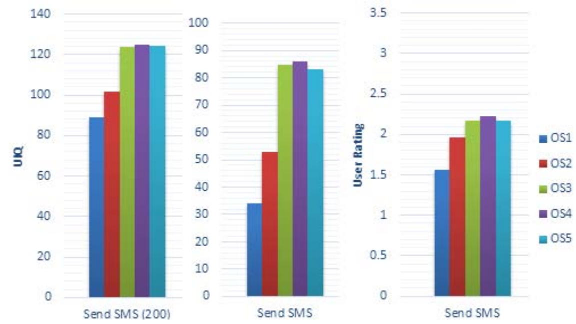

Fig. 9. Values generated for the task of “Send SMS” function.

As it can be seen from the figure, the relative relationship evaluation for different operating systems is consistent with all three evaluations. Survey results indicated that users have found the fourth operating system, OS4 as the “smartest” interface which is easiest to use. OS3 is ranked as the second “smartest”, followed by OS5, OS2 and OS1.

The UIQ calculations revealed measurement results which are very consistent with the survey results. UIQ results should be treated as unit less values that give relative “smartness” values with respect to different systems that are evaluated. In this case, UIQ measurements give exactly the same ranking as the survey results, ranking OS4, OS3, OS5, OS2 and OS1. It is interesting to notice that the relative gap in “smartness” of different operating systems, like difference between OS1 and OS2 found in the survey is correctly reflected in UIQ evaluations.

The results indicate that there is parallelism between UIQ results and the survey results which proves validity of the approach.

REFERENCES

[1] D. Park, J. H. Lee and S. Kim, “Investigating the affective quality of interactivity by motion feedback in mobile touchscreen user interfaces,” International Journal of Human-Computer Studies, vol. 69, no. 12, p. 839-853, Dec. 2011.

[2] W. Haiyan and Y. Baozhu. “A Data-processing Mechanism for Scenario-Based Usability Testing,” Computing, Control and Industrial Engineering (CCIE), vol. 1, pp. 395–398, 2011

[3] Y. Yu and Z. Liu, “Improving the Performance and Usability for Visual Menu Interface on Mobile Computers,” in Proc. of the International Conference on Advanced Visual Interfaces, pp. 369-372, 2010.

[4] H. Liang et al., “A Remote Usability Testing Platform for Mobile Phones,” Computer Science and Automation Engineering (CSAE), pp. 312-316, 2011.

[5] H. Park, B. Kim and K. Lim, "Measuring the Machine Intelligence Quotient (MIQ) of Human-Machine Cooperative Systems," IEEE Trans. on Systems, man and Cybernetics 31:89-96, 2001.

[6] H. Park, B. Kim and G. Lim, "Measuring Machine Intelligence for Human Machine Cooperative Systems using Intelligence Task Graph," in Proc. the International Conference on Intelligent Robots and Systems, 1999, pp. 689-649.

[7] A. Anthony and T. C. Jannett, "Measuring Machine Intelligence of an Agent-Based Distributed Sensor Network System," in Proc. International Joint Conferences on Computer, Information and Systems Sciences and Engineering, Dec 2006.

[8] T. Ozkul, "Determination of Machine Intelligence Quotient (MIQ) Gain of a Mechatronic System by Addition of Visual Guidance," in Proc. International Symposium and Mechatronics and its Applications, Mar. 2009.

[9] El Zarka, Ahmed, and Tarik **Ozku**l. "Assessment Metrics for Intelligence Level of Human-Computer Interface." Journal of Emerging Trends in Computing and Information Sciences 4, no. 3 pp. 319-352 http://cisjournal.org/journalofcomputing/archive/vol4no3/vol4no3_12.pdf