A Framework to Identify Critical Design Parameters for Enhancing User's Satisfaction in Human-AI Interactions

IOP Conf. Series: Journal of Physics: Conf. Series **1284** (2019) 012036 doi:10.1088/1742-6596/1284/1/012036

A Framework to Identify Critical Design Parameters for Enhancing User's Satisfaction in Human-AI Interactions

J G Shin[1,2], J B Kim[1,2] and S H Kim[1,2,*]

[1]BrainKorea21[PLUS] Cultivating Creative Design Engineer Project Group
[2]Graduate school of Industrial Engineering, Kumoh National Institute of Technology, Gumi, Republic of Korea

[*]E-mail: kimsh@kumoh.ac.kr

Abstract. The advancements in artificial intelligence technology have made changes in how people interact with systems. Unique features and user requirements of Human-AI Interactions (HAII) need to be identified with respect to those of Human-Computer interactions (HCI). This study proposes a way to find critical parameters of interaction design for enhancing user's satisfaction when people interact with intelligent systems through voice user interfaces. We summarised distinguished user requirements for intelligent products identified from previous researches. Then match them with design parameters in terms of performance indexes that will make differences in the user's satisfaction. The interaction scenario was set as users ask simple questions with their own voices to the system and the system answer to the questions with synthesized voices after it got to the answer by AI function. The critical performance indexes derived are the number of trials to get the right answer for a question, response time to get to the next interaction, sentence structures of the answer, and pace of the answer. An experimental setup is ready to evaluate user's satisfaction among different levels of the above performance indexes by Wizard of Oz design method applied on a voice user interface we implement. We are going to validate the effects of performance indexes in HAII on the user's satisfaction, which will be measured in terms of verbal and non-verbal measures.

1. Introduction

Through the development of artificial intelligence technology, various intelligent products such as AI speakers (Google's Google Home®, Amazon's Echo®) and virtual assistant systems (Samsung's Bixby, Apple's Siri) on mobile devices are being released [1], resulting in interaction between users and artificial intelligence [2]. Those interactions are called as Advanced Human-Computer Interaction or Human-AI Interaction (HAII). There is a tendency to recognize and understand them as a new version of Human-Computer Interaction (HCI) [3]. But Terry Winograd (2006) argued they differ from the existing Human-Computer Interaction (HCI) [4]. The intelligent products require users to interact with them in new ways. Although HAII has not yet fully established, it is necessary to check whether the user will be satisfactory while interacting with these brand new systems. It is also to check whether there is a need for establishing new criteria to evaluate the goodness of HAII design rather than utilizing traditional way of HCI evaluation method [5]. In this study, a few design parameters that may affect the user's satisfaction are identified by analysing a user's requirements for intelligent products. Finally, we devise and implement an evaluation system for checking out the effect of those design parameters on user's satisfaction while interacting with the intelligent system through the VUI (Voice User Interface).

2. Method

The overall study flow is shown in Figure 1 below.

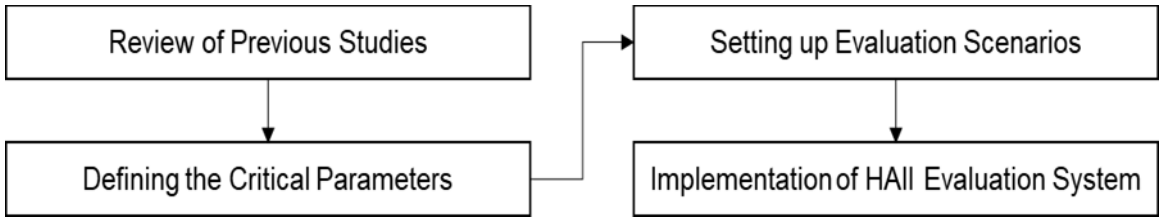

Figure 1. Flowchart for building HAII Evaluation System.

(1) In order to identify a few critical design parameters of intelligent products, user requirements were gathered by reviewing previous studies. Critical design parameters that can affect the satisfaction of intelligent product are drown out by mapping these user requirements with performance indexes. Commercialized intelligent products such as Samsung's Bixby or Apple's Siri were investigated as the source of variations for the design parameters.

(2) A model HAII system was designed and implemented to conduct experiments for evaluating the effects of identified design parameters. Since it is not easy to control the design parameters of the commercialized products, we took the WoZ (Wizard of Oz) evaluation method. It means that this model HAII system was designed as it can provide different interactions with modified design parameters in random order by researchers in the rear side, but the subjects are feeling and believing that they interact with the real AI system [6].

3. Result

3.1. Definitions for Critical Design Parameters

We collected user requirements in two previous studies related to user requirements for intelligent products. User requirements are shown in Table 1 below.

Table 1. User Requirements for HAII.

Paper	User Requirement	Selection
Laila Dybkjær and Wolfgang Minker 2008. [7]	• Modality appropriateness	
	• Input Recognition adequacy	
	• Naturalness of user speech relative to the task including coverage of user vocabulary and Grammar	
	• Output voice quality	V
	• Output phrasing adequacy	
	• Feedback adequacy	V
	• Adequacy of dialogue initiative relative to the task	V
	• Naturalness of the dialogue structure relative to the task	
	• Sufficiency of task and domain coverage	
	• Sufficiency of the system’s reasoning capabilities	
	• Sufficiency of interaction guidance	V
	• Error handing adequacy	
	• Sufficiency of adaptation to user differences	

	• Number of interaction problems • User satisfaction	
Valéria Farinazzo et al, 2010. [8]	• Appropriate modality • Suitable Feedback • User diversity and user perception • Appropriate phases out • Output Voice Quality • Proper entry recognition • Appropriate dialog start out and adequate instruction about how to interact with the application • Help tool • Error Prevention • Handling errors	V V V V

With these chosen user requirements, we analyse the functional aspects of the intelligent systems in terms of design parameters so that we can map the user requirement and the design parameters. Eight user requirements were selected and then mapped with four design parameters as shown in Table 2 below.

Table 2. Critical Design Parameters Identified.

User Requirements	Critical Design Parameters
• Appropriate dialog start out	• Response time to get to next interaction
• Error Prevention • Error handling adequacy	• Number of trials to get the right answer for a question
• Output Voice Quality	• Pace of the answer
• Feedback Adequacy • Adequacy of dialogue initiative relative to the task • Naturalness of the dialogue structure relative to the task • Output phrasing adequacy	• Sentence structures of the answer

The four design parameters derived from user requirements are defined as follows:

- Response time to get to next interaction - The time it takes for an intelligent product to respond to a user's question.
- Number of trials to get the right answer for a question - The number of the trial that an intelligent product challenged to give the right answer.
- Pace of the answer - The pace of answers that users can easily recognize.
- Sentence structures of the answer - The amount of information in the answer presented by the intelligent product after the user requests the information.

In order to design satisfactory HAII, an evaluation system is to develop for identifying the effect and proper values of the design variables. Before starting the evaluation experiments we have to specify a range and levels for each design parameter considered as independent variables. On the basis of commercial system values and the previous studies, the range of design parameters to be tested are determined as shown in Table 3 below.

Table 3. Scope of Critical Design Parameters.

Critical Design Parameters	Scope
• Response time to get to next interaction	• Between 1 to 5 sec
• Number of trials to get the right answer for a question	• 1 trials • 2 trials • 3 trials
• Pace of the answer	• 4 Syllable/sec • 6 Syllable/sec • 8 Syllable/sec
• Sentence structures of the answer	• Answer only • Repeat Question then Answer • Repeat Question then Answer with Clear Reference Source

The response time to get to next interaction is 2 seconds on average for the released intelligent product (Bixby&Siri), and a range of 1 to 5 seconds was conceived as a good control group to test its effect on the user satisfaction. Regarding the number of trials to get the right answer for a question, based on a result from a study that VUI is excusable for up to two errors [9], three levels from first to third trials were considered as a good range for the test. Pace of the answer was set at 6 Syllable/sec, which is the average response speed of Bixby as middle levels and set 4 Syllable/sec and 8 Syllable/sec as slow and fast paces. Lastly, the sentence structure of the answer was set as three different cases: feeds the answer back only, repeat the given question before the answer to make the user sure that the system understood the question correctly, and repeat the given questions before the answer with clear and convincing reference source.

3.2. Design for HAII Evaluation System

The HAII evaluation system was designed and implemented so that the system can emulate varying interaction conditions by combining different design parameter values described in the previous section. The system was devised to be good at conducting mixed within-subject design since the time and cost of the testing experiments as well as the fatigue of subjects will be burdensome if all the design parameters are considered as within-subject variables. Response time to get to next interaction and pace of the answer were considered as between-subject variables, and the remaining two parameters (number of trials to get the right answer for a question, sentence structures of the answer) were considered as within-subject variables. By this design, any subject can finish their tasks for only 9 trials. As an experimental session begins, the subject is given a question represented on a computer screen. The subject needs to initiate the interaction by asking the question to the HAII with their own voices. Then the HAII system feeds the answer back depends on the experimental protocol. The overall flow of the interaction scenario is shown in Figure 2 below.

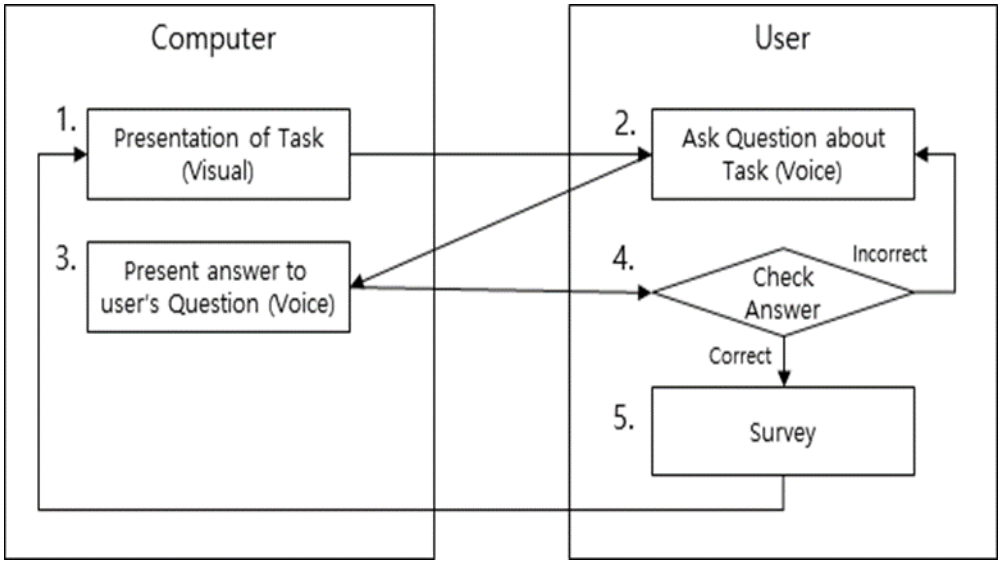

Figure 2. Flowchart for HAII Scenario.

Figure 3 illustrates the different scenarios or trials of experiments that a subject needs to get through for a given question.

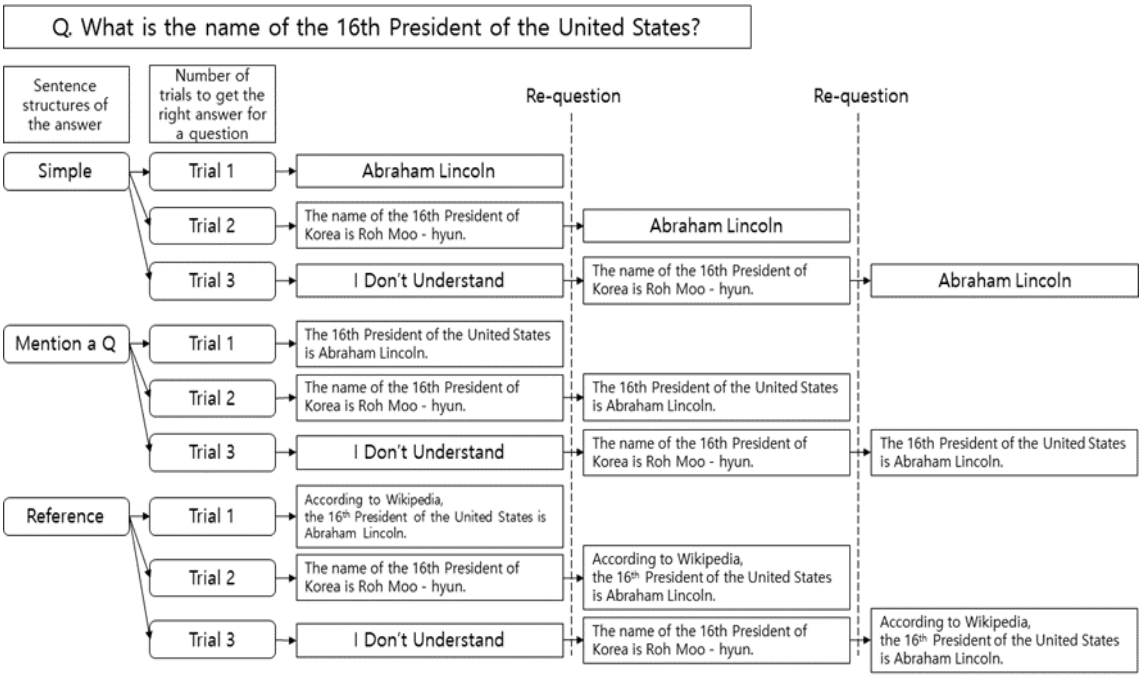

Figure 3. Example: Answer Design for Task.

In the given example above, a subject is given a question that "Who is the 16th President of the United States?" The subject asks to the evaluation system through the VUI for searching the answer for the question, and the evaluation system randomly feeds one of the nine scenarios back. The interface of the HAII evaluation system is shown in Figure 4 below.

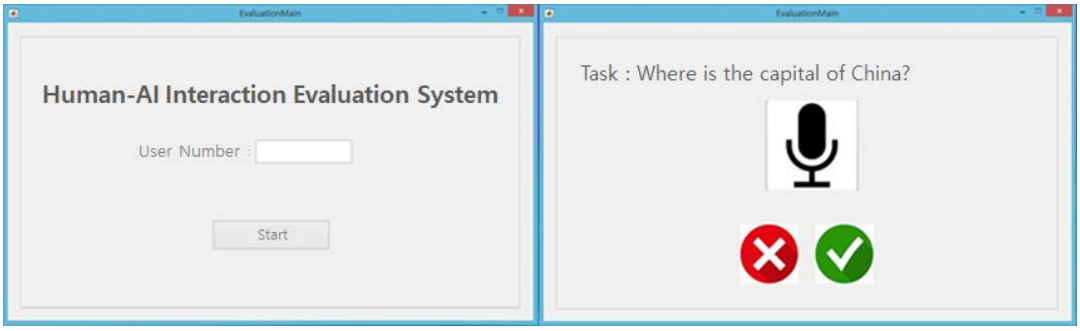

Figure 4. HAII Evaluation System Interface.
(Left: Start Screen, Right: Evaluation Screen)

4. Discussion

In this study, we derived four critical design parameters of the HAII that can affect user satisfaction during interacting under VUI environment. These design parameters were carefully identified by mapping the selected user requirements on functional performance indexes of HAII. We focused on 8 out of 25 user requirements that were dealt in previous studies. We took the 8 by eliminating those ones that are redundant or does not directly related to the quality of answers from the system. To get more thorough information about how to design satisfactory HAII, it is supposed to collect the user requirements more diversely in further studies including direct surveys on experienced users of the system.

The current version of HAII evaluation system implemented in this study considered the 4 design parameters with specified ranges, but the system is devised as flexible for the changes of parameters as well as their ranges. Since we adapt the Wizard of Oz method, it is easy to add or remove a set of design parameters having a variety of ranges. The evaluation system was developed as an embedded software so that it has another advantage of portability makes it free to apply on a variety of environments using VUI.

To validate the criticality of the four parameters in the HAII design, it is mandatory to proceed with further studies regarding real evaluation of user satisfaction. User satisfaction is a subjective matter and that is used to measure through verbal responses by rating scales or semantic differential method [10]. Non-verbal evaluation protocols utilizing bio-signals such as EEG, heart rate, facial expressions are suggested recently as alternative ways to supplement the verbal protocols [11]. In the next study, we are going to utilize both protocols to find the effects and the proper values of the four parameters derived in this study.

Acknowledgments

This research was supported by the Brain Korea 21 Plus Project (Dept. of Industrial Engineering, Kumoh National Institute of Technology).

References

[1] Gustavo L, Luis Q and Luis A G 2017 *Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-based Natural User Interfaces*. (Int. Conf. on Applied Human Factors and Ergonomics) pp 241-250
[2] Jinbae K, Saleh S A A S and Hyunsoo L 2018 *Pattern Recognition and Classifier Design of Bio Signals collected in HAiI* (Proc. of KIIS Fall Conf. Vol 28 Issue 3) pp 9-10
[3] Johnson H *Relationship between User Models in HCI and AI* 1994 (IEE Proc. Computer Digital Techniques Vol 141 No 2)
[4] Terry W 2006 *Shifting Viewpoints: Artificial Intelligence and Human–Computer Interaction* (Artificial Intelligence Vol 170 Issue 18) pp 1256-1258
[5] Jonggyu S, Ingwon J and Sangho K 2018 *A Study on the Difference between Human-AI*

Interaction and Human-Computer Interaction (Proc. of Ergonomics Society of Korea Fall Conf.)
[6] Bella M and Hanington B 2012 *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions* chapter 99 p 204
[7] Laila D and Wolfgang M 2008 *Recent Trends in Discourse and Dialogue* (Berlin: Springer-Verlag)
[8] Valeria F, Martins S, Andre L S K and Joao S O N 2010 *An Empirical Approach for the Evaluation of Voice User Interfaces (User Interfaces)* ed Rita M (Hungary: Eötvös Loránd University) Chapter 10
[9] Myeongji L and Kihyung H 2011 *Design and Implementation of a Usability Testing Tool for User-oriented Design of Command-and-Control Voice User Interfaces* (Phonetics and Speech Sciences Vol 3 No 2) pp 79-87
[10] Sangho K, Sunah K, Jongkyu S and Jeongyoon A 2016 *A Human Sensibility Ergonomic Design for Developing Esthetically and Emotionally Affecting Glass Panels of Changing Colors* (Journal of the Ergonomics Society of Korea Vol 35 Issue 6) pp 535-550
[11] Dongmin S, Dongil S and Dongkyoo S 2015 *Research of Real-Time Emotion Recognition Interface Using Multiple Physiological Signals of EEG and ECG* (Journal of Korea Game Society Vol 15 Issue 2) pp 105-114