Assignment M4:

Evaluations for FaceID Redesign Prototypes

Cleo Zhang

yzhang3761@gatech.edu

***Abstract*— This report evaluates the FaceID redesign prototypes proposed in M3. First, we will be qualitatively evaluating the "Indoors Shopping" mode, empirically evaluating the orientation detection prototype, and predictively evaluating the voice controlling, then we state an execution plan for the next assignment.**

1 QUALITATIVE EVALUATION: INDOORS SHOPPING MODE

This section will use the think-aloud method to evaluate the indoor shopping prototype, as the indoor shopping mode was verbally stated in M3 and think-aloud usually works well for verbal prototypes.

1.1 Evaluation plan

The participants for this evaluation will be recruited through friends and fellow students in CS 6750 via my social media and the Ed discussion board. The participants are required to be iPhone users.

The whole process will occur online with Zoom calls. The participants will be provided with a verbal description of the Indoor Shopping mode first, then a context to think aloud how to use the indoor shopping prototype to complete a contactless payment. The calls will be recorded by the built-in recording tool in Zoom to capture the evaluation results.

1.2 Evaluation content

We will be primarily applying the think-aloud protocol to gather feedback by asking the right questions. Since research shows that the way the participants interact with the interface may change when they are asked to think aloud, we

may want to mix the post-evaluation protocol to eliminate this bias from think-aloud.

1.2.1 *Directions*

- The ultimate goal: to complete a contactless payment.
- Context:
    - Shopping in a grocery store;
    - Masks are mandatory in the store;
    - FaceID is enabled on iPhone, but faces cannot be recognized because of the masks.

1.2.2 *Questions*

Below questions will be asked during participants interfacing with the indoor shopping mode.

1. How easy is the indoor shopping mode to be found?
2. After you have found the mode settings, do you know what to set?
3. How often do you think you would forget to set this mode before shopping? 0% of the time, 20% of the time, 40% of the time, 60% of the time, 80% of the time or 100% of the time.
4. What do you like about this interface?
5. What do you not like about this interface?
6. Overall how satisfied are you with this interface? On a scale of 0(not satisfied at all) - 10(very satisfied).
7. Do you think you spend longer or shorter time completing a contactless payment compared to using FaceID?

1.3 Addressing the requirements

This evaluation addressed the below requirements listed in M2:

- *Such a feature should be novices friendly and easy to be picked up* - by collecting the answers for the first and second questions above, we would understand if this indoor shopping design is novice-friendly or not.

- *The overall time spent on contactless payment should not be longer than directly typing in passwords to pay* - this requirement is addressed by the last question above, and we would understand if this requirement is met by examining the answers collected from this question.

With the data inventory stated in M1 and the answers to the above questions from participants, we can evaluate the "indoor shopping" design from functionality, usability, learnability and accessibility.

2 EMPIRICAL EVALUATION

This section will examine the Orientation Detection prototype from M3 by applying empirical evaluation. We will first discuss the control/experimental conditions, then null and alternative hypotheses, followed by the experimental method used and possible lurking variables.

2.1 Control and experimental conditions

What are being tested - three primary metrics will be tested for this prototype: (1) user's ability to navigate between FaceID and password typing; (2) the learning curve of this design; (3) the effectiveness of using this design to complete a contactless payment.

What is the point of comparison - the point of comparison will be users' opinions about the overall effectiveness of this prototype. We will be comparing the experience of this prototype with the existing FaceID design.

2.2 Null and alternative hypothesis

The null hypothesis - the null hypothesis is expected to be true if the alternative hypothesis cannot be proved. In this case, the null hypothesis is that the orientation detection prototype is as efficient as the existing FaceID design.

The alternative hypothesis - the alternative hypothesis is what we expect to prove true on our prototype. Usually, we want to have numeric results from the control and experimental groups and draw conclusions by seeing significant differences

between the two. The alternative hypothesis in this evaluation is: the orientation detection prototype acts more effectively than the current FaceID design.

2.3 Experimental method that will be used

We will be using the within-subjects experiment for this evaluation. The treatments here are using the existing design of FaceID and the orientation detection prototype to complete a contactless payment at grocery stores.

Participants will be randomly assigned to treatment groups, and each participant will experience both treatments. Plus, we also want to randomly assign the participants to what order they will receive the treatment, as the order of which treatment that a participant first sees may affect the experiment results.

The participants will perform the same payment task using the existing FaceID design and the orientation detection prototype as their conditions. We will be monitoring the time they spend on each interface to complete a contactless payment as our experiment results.

We will be collecting and analyzing the data generated by participants for each treatment in our experiment. The goal is to see if there is a significant difference between the results from the two treatments, and we want to draw an objective conclusion based on the statistics we have. If the average time spent on the orientation detection prototype is less than the existing design, we can prove the alternative hypothesis true.

2.4 Lurking variables

The first confounding variable is the order of which treatment that a participant first sees. We have discussed how to eliminate the influence from this variable in the previous section.

The second variable is the physical shopping environment. Even for the same grocery store, it would be tough for us to make sure each payment behaviour happens under the same context - how busy the grocery store is, the internet speed, and how fast cashiers react to the payment request.

3 PREDICTIVE EVALUATION

Predictive evaluation is an evaluation method we can use without involving real users. We apply predictive evaluation to the voice controlling prototype because both qualitative and empirical evaluation is not applicable in this case - think-aloud may confuse the device about what verbal command it should take as an input, and the cost of the empirical evaluation is too much.

In this section, we will analyze and compare the voice controlling prototype with the current design of FaceID by constructing a GOMS model. And we use the requirements defined in M2 as the selection rules to evaluate our prototype.

Initially, users have all products scanned and are ready to pay for the order, and the ultimate goal is to complete a contactless transaction at a grocery store. Below are the specific methods and the operators that users need to perform to get the job done.

The current FaceID design - put the device close to the terminal → seeing the Face detection console → adjust the face to fit in the camera → face recognition failed due to the mask → prompt asking if you want to type in password instead → click yes → type in password while holding the device close to the terminal → transaction approved.

The voice controlling prototype - verbally speaks out the command "Hi Siri, use the password for now." → hold the device to the terminal → the keyboard pops up → type in the password → transaction approved.

By seeing the above flows, it is clear that the voice controlling prototype bridges the gulf of execution by simplifying the tasks and subtasks. However, the voice controlling feature can be easily affected by the external environment. For example, what if the grocery store is noisy and the device does not know what verbal command it should take as a valid command. Such flaws will need a longer time to be fully fixed with voice controlling technology evolving.

4 PREPARING TO EXECUTE

We will be choosing the below two evaluations for the following assignment:

- An empirical evaluation for the orientation detection prototype.
- Qualitative evaluation for the indoor shopping prototype.

The reason is that the execution plans are applicable for those two, but it would be hard to mimic all the factors that may affect the user experience for the voice controlling prototype and draw an objective conclusion. Given the limited time and resources, I believe the empirical evaluation for the orientation detection prototype and qualitative evaluation for the indoor shopping prototype can give us more valid experiment results.