

Review

This paper presents a user interface for [REDACTED]. The private space and the shared space are visualized by [REDACTED]. The user touch and hold to [REDACTED] between this space. A user study is presented to compare the effect of [REDACTED] to the usability [REDACTED].

Overall

This paper investigates an appealing domain of [REDACTED] in mobile devices. Nevertheless it falls short in defining a compelling research agenda, surveying the related work, justifying the interaction design, establishing the internal and external validity of the study. Therefore, I could not recommend accepting this paper.

Review Criteria

Novelty: [*]
Methodology: [*]
Presentation: [***]

Legend

(+) positive aspects
(-) negative aspects
(*) suggestions
The number of symbols in the parentheses indicates the magnitude of severity / virtue.

Content

BACKGROUND AND RESEARCH AGENDA

(+) Interesting domain

Although the interaction in [REDACTED] is not new, using it in the context of mobile device raises many issues, e.g., [REDACTED]. I applaud the brevity of the authors to investigate in this domain.

(--) Why do we need another interaction technique for [REDACTED]? Presently, we already have many techniques ranged from [REDACTED] (standard iOS [REDACTED]; [REDACTED] (CHI '04)), [REDACTED] ([REDACTED], AVI '04'), [REDACTED]. The author should show that they are aware of these techniques and identify the "gap" in the usage scenario to make the proposed interaction technique more compelling.

On the other hand, if the [REDACTED] is used as a scenario to test the interaction with [REDACTED], the authors should clarify upfront.

(-) Missing relevant literature:

I recommend reviewing the literature as stated above and in the following comments to create a strong "wall" to back the research. Reference 1, 3, and 4 could be omitted in a short paper submission because they were used only to support general knowledge.

INTERACTION DESIGN

(--) What is the additional benefit of the [REDACTED] for this task? The use case presented is a [REDACTED] which needs only two separated space (private and shared). What are reasons against splitting the screen to explicitly separate the private and the shared area? While screen splitting occupies the same screen real estate as it currently used in the proposed UI, it has a benefit of a clear distinction between the areas without [REDACTED].

USER STUDY

(--) Incorrect methodology for testing H1:

While the aim of H1 was to assess whether the user understand the interaction technique, the study confounds this by training the user to use the program. Since the interaction technique is straightforward [REDACTED] training would guarantee that all participants understood the concept. Hence, H1 was not tested.

To test whether the user can understand the interaction design, the experimenters should let the user discover the user interface by themselves and collect data to answer the following questions: Did they find the gesture? What other gestures did they try? How long does it take to learn the gesture? What is the mental model that they have at first sight of the UI? How does the mental model changes as the user try the system? Think aloud or constructive interaction method can be used for this purpose.

(--) H2 study is not necessary:

It is known that [REDACTED] leads to a better perception of [REDACTED] ([REDACTED], MoViD '10). The same paper also ranked the importance of different depth cues.

(--) H2 study lacks internal validity:

The choices of the cues presented in this study are also limited to [REDACTED]. Therefore, H2 should be rephrased to investigate subjective preference of [REDACTED].

The result of the study might reflect the quality of the [redacted] implementation rather than [redacted]. The author should include the image of the [redacted] implementation to support this, and should justify the choices of the visual cues.

If the author want to investigate whether the [redacted] convey the meaning of private or shared space, the users should not be trained to use the interface before hand, and the test should be separated from the [redacted].

(-) The reported rating (Fig. 3) is not meaningful: Since the test was done in single condition, there is no comparative measurement, i.e., a gold standard, to calibrate the result. User's opinion can be biased to in favor of the system that the experimenter implemented.

Also, used questionnaire is not standardized. I suggest the authors to have a look at standard usability questionnaires, e.g., PSSUQ or SUS, for future studies.

(-) Rating results in Fig. 6 should not be simplified to three comparison categories. The 5-point score should be analyzed with a statistical method that supports within-subject non-parametric data, e.g., Friedman's ANOVA.

References

[redacted]

[redacted]

[redacted]

[redacted]

Formatting

- (-) What are the error bars in Figure 3 and 5? Range, SD, SE, or 95% CI?
- (-) The font in reference numbering is not consistent with the rest of the text.
- (-) References are not consistently formatted, e.g., "Proceedings" vs. "Proc."

Overall Rating

===

1 - Strong reject

Inspiration

===

1 - Uninspiring

Expertise

===

3 - Knowledgable

Rate your review

===

3 - Pretty OK

This paper compares three [redacted]. They were evaluated with three measures: (1) perceived effectiveness, (2) perceived usefulness, and (3) information retention. Two between-subject user studies show that [redacted] (N=45, 21).

Overall

I appreciate that the authors tap into the relatively unexplored area of [redacted] with thorough studies and the analyses. Nevertheless, three crucial issues seriously hinder me to argue for accepting this paper: (1) The benefit of [redacted] (2) The [redacted] confounded the study, and (3) The severely limited generalizability of the result.

[redacted] Review Criteria

Originality: [**]
Methodology: [**]
Presentation: [*****]

Legend

(+) positive aspects
(-) negative aspects
(*) suggestions

The number of symbols in the parentheses indicates the magnitude of severity / virtue.

Content

(+) Extensive literature review: I appreciate the extensive literature review on [redacted]

RATIONALE OF THE STUDY

(--) Why [redacted]? I could not find any justification of [redacted]. Since [redacted]

[redacted] [9], why not u [redacted]

?

(-) Weak justification of [redacted]: Although the authors appealed to [5] and their experience, the connection between [redacted] presented in [5] ([redacted]) and the [redacted] used in the paper ([redacted]) were unclear. Could this leave other possible [redacted] unexplored? The authors may need to address this limitation in their discussion.

(*) A formative evaluation would be more useful than the summative evaluation. It would allow others to follow the step to construct the similar [redacted]. cf. [28]

CONFOUNDING VARIABLES

(---) Three [redacted] differ in the [redacted]. This is apparent when focusing only on the part that is distinct among the three [redacted]. While the [redacted] design lacks the [redacted], [redacted] design had a lot of additional information about [redacted]. Consequently, the [redacted] complexity of the three designs was extremely deviated.

The differences observed in the result could be influenced by the difference in the [redacted] complexity. This was also confirmed by the qualitative responses, time, and effort to read as reported in 5.1.1.

(--) [redacted] advice confounds the usefulness rating: The [redacted] advice is commonly provided at the bottom of all three designs in the same format. Since the study was between-subject design, and the participants were asked to rate the usefulness of [redacted], they could have judged the usefulness based on the [redacted]. In particular, it is unfair that the [redacted]

(-) The [redacted] could also confound the test. Nevertheless, the effect could not be stronger than the two factors as aforementioned.

GENERALIZABILITY

(---) Since each metaphor was represented only by one [redacted], and since the [redacted] could be extremely influenced by [redacted], I doubt whether the results presented in this paper will be generalizable to the other [redacted] using the [redacted]. A future work to prove the superiority of one metaphor than another would be to [redacted]. The similar test had been done in [17].

(*) The finding indicating the differences in the participants' background knowledge among [redacted] is useful, but it is based on the limited population. Replicating the experiment with an online survey would give a better representation of the general population.

EVALUATION

(+) Strong evaluation procedure, participant recruitment, and statistical analysis

(-) In my opinion, the second user study is just the fourth condition of the first user study. Since the authors used between-subject design, there is no point of separating them.

(*) The authors mentioned the purpose of the second study is to test the [REDACTED] This was not mentioned elsewhere.

(*) Please clarify whether each of the analyzed scores (effectiveness and usefulness) was summarized from multiple questions or a single question. (The author mentioned [REDACTED] questions in the questionnaire with reversed directions.)

(*) Results from Kruskal-Wallis tests were reported with incorrect degree of freedom (3). Nevertheless, the p-values were correctly calculated based on the correct degree of freedom (2 = 3 conditions - 1 summarized estimate).

(*) Retention comparison: I wonder why the authors chose pre-study results instead of post-study results to compare with the one-week result.

Formatting

The paper was written in a good language style, and the graphics were in the appropriate quality.

Postscript

It is hard to evaluate [REDACTED] comparatively, especially when the [REDACTED] is publicly known (cf. [16] in which the participants were tested with the knowledge that is unknown to them beforehand). Nevertheless, the authors did a good job of reporting differences in users' background knowledge and their post-test attitude on their [REDACTED].

Unfortunately, the big elephant in the room was the rationale, the confounding, and the generalizability that I mentioned above. Despite being a well-conducted research, I could not bring myself to vote for acceptance of this paper.

Lastly, I am not confident that this paper will fit the scope and interests of the [REDACTED] audience. Nevertheless, I leave this judgement to the ACs, and this factor did not influence my score.