

User Sketches: A Quick, Inexpensive, and Effective way to Elicit More Reflective User Feedback

Maryam Tohidi

University of Toronto
maryam.tohidi@gmail.com

William Buxton

Microsoft Research
bibuxton@microsoft.com

Ronald Baecker

University of Toronto
rmb@kmdi.utoronto.ca

Abigail Sellen

Microsoft Research
asellen@microsoft.com

ABSTRACT

Our aim is to introduce techniques that allow for active involvement of users throughout the design process, starting with the very early stages of ideation and exploration. The approach discussed in this study augments conventional usability testing with a *user sketching* component. We found that enabling users to sketch their ideas facilitated reflection, and provided a rich medium for discovery and communication of design ideas. We believe that this technique has the potential to complement usability testing in general, in order to generate “reflective” as opposed to purely “reactive” user feedback.

Author Keywords

Sketch, Design, Prototyping, Usability Testing, Methods, User Centered Design.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

The main aim of our research is to enable end users to contribute at every stage of the design process, including the early stages of ideation and exploration. In these early phases the focus is on getting the *right design*, as opposed to getting the *design right*. Therefore generating diverse and creative ideas and exploring a variety of alternative solutions is the main goal here.

Usability testing (UT) is well-known and commonly used to involve users in various stages of the design process. Think-aloud techniques, questionnaires, interviews, and the observation of behavior during task performance are common instruments for UT. While these techniques are good in eliciting comments and criticisms and identifying problems, they fall short in facilitating reflection which is required to generate design ideas and alternative solutions.

It has been suggested that current UT practices tend to elicit more “reactive”, rather than “reflective” user feedback [9]. This has been confirmed in [12] which showed that UT participants generate significantly more reactive *comments* and *criticisms* than reflective *design suggestions* or *redesign proposals*.

Yet, it has been shown [4] that developers prefer to receive reflective *redesign proposals* as opposed to *usability problems* as an outcome of UT. What [4] fails to address is how to encourage, or better yet, enable end users to generate more such redesign proposals.

Smith *et al.* [9] suggest that in order to receive more reflective feedback “participants need to engage in a deeper level of reflection and projection”, and suggest a multi-session approach to UT that is much longer and more expensive. Unfortunately high cost is a limiting factor in the use of UT technique. Cost pressures have already led to the adoption of “discount” techniques. While proving useful, some discount techniques such as Heuristic Evaluation [5] do not involve end users. The objective of our work is to provide the right means for generating reflective feedback, without adding significantly to the cost. Furthermore, we want to achieve this without eliminating the active involvement of users in the process. The proposed technique is *user sketching*.

In this paper we provide evidence from a study that shows how user sketching can elicit reflective feedback that is complementary to that which is obtained using conventional techniques (e.g. questionnaires, interviews, and think aloud protocols) and which might otherwise be missed. The technique also provides reactive data that is comparable to that obtained with those same conventional techniques. Furthermore, it does that at a fraction of the cost of videotape analysis, for example. Since the user sketches are quick to create and take only a brief time to analyze, and provide reactive as well as reflective feedback, we believe that the user sketching technique is a great addition to our current repertoire of UT techniques.

The Language of Design

Coming up with design solutions is a highly reflective process, and requires a special medium that facilitates this type of reflection. In “The Reflective Practitioner” Schon [7] speaks of Quist, an architect who demonstrates reflection-in-action:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NordiCHI 2006: Changing Roles, 14-18 October 2006, Oslo, Norway

Copyright 2006 ACM ISBN 1-59593-325-5/06/0010...\$5.00

As Quist reflects on the unexpected consequences and implications of his moves, he listens to the situation's back talk, forming new appreciations which guide his further moves... Quist makes his moves in a language of designing which combines drawing and speaking... Drawing and talking are parallel ways of designing, and together make up what I will call the *language of designing*.

Since it is fundamental to our approach, let us begin with a brief discussion of the role of drawing/sketching in design.

A Sketch of Sketching

Sketching is fundamental to ideation and design. Traditional disciplines such as industrial design, graphic design and architecture make extensive use of sketches to develop, explore, communicate and evaluate ideas. In user interface and interaction design, practices such as "Wizard of Oz" techniques [3] and paper prototyping [10] can perform the role of sketching. While being quick and inexpensive, such "sketches" produce comparable usability testing results as their more expensive, high-fidelity alternatives [1, 8, 14, 16].

Sketches allow for a dialog between the sketch and the viewer (even if the viewer is the sketcher himself) that facilitates better understanding of the problem and in turn generation of new ideas:

Designers do not draw sketches to externally represent ideas that are already consolidated in their minds. Rather, they draw sketches to try out ideas, usually vague and uncertain ones. By examining the externalizations, designers can spot problems they may not have anticipated. More than that, they can see new features and relations among elements that they have drawn, ones not intended in the original sketch. These unintended discoveries promote new ideas and refine current ones. This process is iterative as design progresses. [11]

Figure 1 represents the conversation between sketch and human. The sketch is created from current knowledge (left arrow) in a form of "seeing that" reasoning [2]. "Reading" or interpreting the sketch (right arrow) generates new knowledge in a form of "seeing as" reasoning [2]. The iterative nature of this process allows for an increasing improvement of the sketch/design even if the person making the sketch is the same person that is "reading" it.

In conventional UT, the interaction designer communicates with the user largely by means of an "interactive sketch", such as a paper prototype. Users then explore the "sketch"; however, when it comes to the user communicating back to the designer, the modalities used are mainly spoken or written language, *via* "thinking-aloud", interviews or questionnaires.

The concept that we explore in this study is this: just as we communicate our design ideas to users using more than just words (i.e., interactive sketches), why not likewise provide users with a richer medium for communicating their feedback to us?

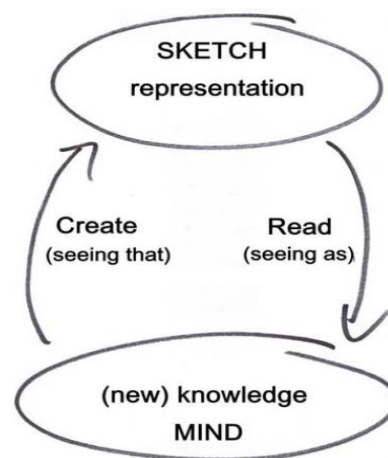


Figure 1: A sketch of a dialog with a sketch

Overview of Our Work

Our work has taken place in two stages. In our first study [12], we examined the differences that would result between a usability test that exposed users to a single design, versus one where they tried three different alternatives. The "product" was a touch-sensitive House Climate Control System (HCCS). Three distinctive prototypes were developed, each reflecting a different design language: Circular (Figure 2), Tabular (Figure 3), and Linear (Figure 4). All three were consistent in terms of fidelity, functionality and quality.

Aside from incorporating more than one design, the techniques used (low-fidelity prototyping, task-based user testing, think-aloud, questionnaires and interviews) were common practice taught in standard texts such as [5, 6, 10].

Our second study is the topic of this paper. At the end of the first study we provided subjects with a blank piece of paper and ask them to sketch their ideal system design. We said that a quick drawing that shows the core features of the main interface would be fine. Since this was the last part of the experiment, no time restrictions were imposed. Here we compare the results that we got from the first study with what we learned from a "reading" of the user sketches.

WHAT VERBAL FEEDBACK TOLD US

This section summarizes our earlier study, but only inasmuch as it supports the discussion of our new results. For more details, the reader is referred to [12].

48 participants were divided into four groups of 12, three of which were presented with only one of the three prototypes, and the fourth sequentially presented with all three. Subjects performed a standard type of usability test, consisting of a think-aloud protocol while performing a set

of pre-defined tasks, followed by a questionnaire and a semi-structured interview. The latter aimed to allow participant to verbalize their likes and dislikes of the design (i.e. reactive feedback) and make suggestions for design improvement (i.e. reflective redesign proposals).

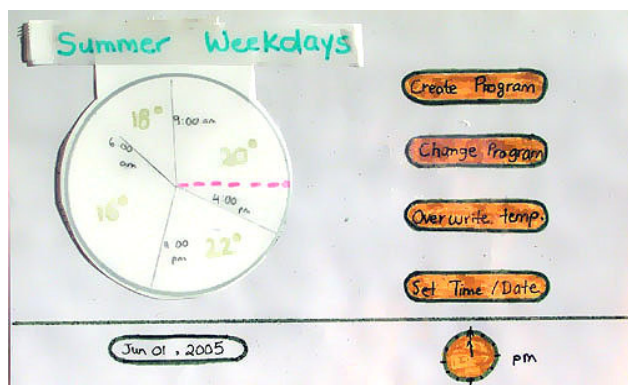


Figure 2: The “Circular” paper prototype

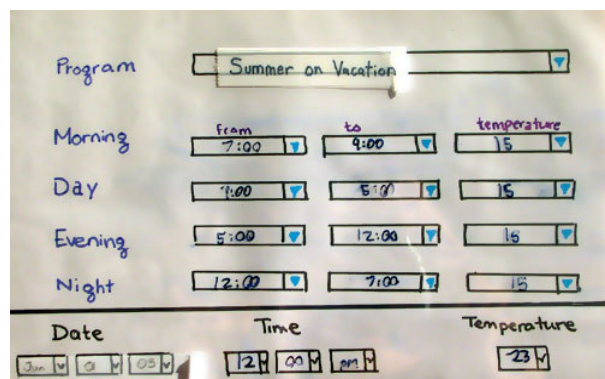


Figure 3: The “Tabular” paper prototype

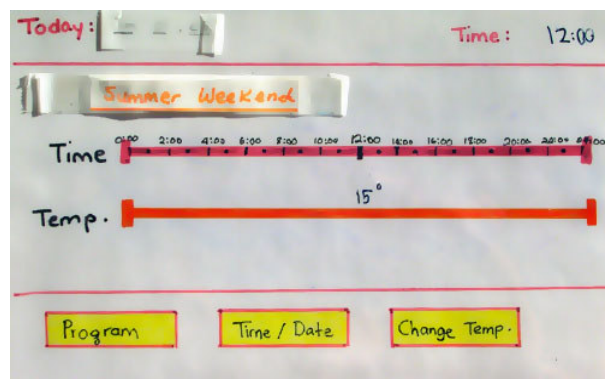


Figure 4: The “Linear” paper prototype

User Ratings of the Designs

Our first hypothesis was that designs would be rated higher when seen alone than they would when seen in combination with other designs. We found this to be the case:

Within each questionnaire the prototype(s) viewed were rated on 12 measures: four ease-of-use ratings, seven design

Paper Prototype	Circular N = 12	Tabular N = 12	Linear N = 12
Seen alone	9.00 (SD = 0.90)	8.67 (SD = 0.79)	8.08 (SD = 0.90)
Seen in a Group of 3	8.08 (SD = 2.15)	8.08 (SD = 1.73)	6.50 (SD = 1.83)

Table 1: Mean and (standard deviation) of overall ratings for the three designs either seen alone or within a group.

dimensions and an overall rating. The results show that, on the basis of overall ratings (see Table 1), prototypes seen in isolation were rated higher than those seen in a group of alternatives. In addition, the Linear design was significantly the worst of the three, while the Circular and the Tabular variations were more comparable to one another. (For more detailed analysis, see [12].)

Feedback from the Think aloud and Interview Data

Another hypothesis was that users who were exposed to all three designs would give us more redesign proposals than those who saw only one. We were surprised that this turned out *not* to be the case. Subjects' think aloud protocols, post-study interviews and questionnaires were all dominated by reactive rather than reflective feedback. In other words, subjects' verbal feedback was more likely to consist of comments on or critiques of the designs they had seen rather than suggestions for how they could have been improved.

The think-aloud protocol and interview - the instruments for verbal feedback - were both analyzed from video recordings of the study. Due to the high costs of video analysis only half of the 48 tapes were selected for in-depth analysis: six chosen at random from each of the four conditions. The results of the 24 tapes are summarized in Figure 5. Overall, subjects provided significantly fewer design suggestions than comments and criticisms. This pattern was observed in both the single and multiple prototype conditions, which was not what we expected. In Figure 5 each bar represents the sum of ideas in both the single and multiple conditions.

When asked “What would they change about the design?” participants frequently pointed out the problems but then indicated that they could not say how they would change it. They would say things like, “the morning, day, evening, night - it's maybe a little too repetitive. I don't know. It's like evening, night, its kind of... maybe... or... I don't know! I haven't figured it out”, “I think the main problem is the time blocking. How to change it? Maybe one way is to change this box? I'm trying to think of creative ways to solve this problem. [long pause] maybe it's not a problem for others, just a problem for me. I don't know!”, or “Well I don't know how this would be done, but maybe... I don't know, just the number of interfaces is a little bit daunting”.

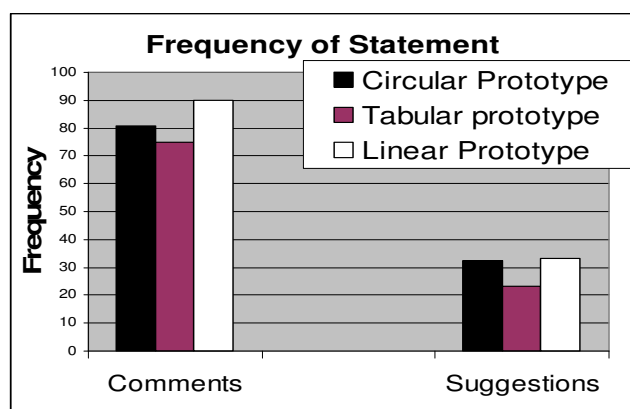


Figure 5: Frequency of statements

Even more frequently participants took back their suggestions shortly after making them, such as in, “I think it might be better to have a 24 hour system. So, I mean, from 17 to 19... or, maybe not!”, “I guess the fact that you can, like, in the first one you can pick how many... oh I guess you can only have four intervals as well... Never mind my idea was shot because I was thinking well what if you want more intervals then that”, or “The time is kind of small I would say; specially for older people, maybe they might have trouble seeing, I guess. [later in the interview] I think the time and date ... I guess - to me it's a side thing. I mean, like, if you are looking for the time, like, you wouldn't look at it (a thermostat) - that's what I mean. You'd look at a clock...”

The data, as reflected in these quotes, highlights the participants' lack of confidence in their own redesign suggestions. It appears that the process of thinking and verbalizing one's thoughts at the same time is very distracting: participants tend to speak in broken sentences, and seem to struggle between problem solving, verbalizing ideas, and reflecting on previously mentioned ideas.

WHAT SKETCHING TOLD US

This brings us to our second, follow-on study. After subjects were “finished” the UT reported in [12], we asked them: “Please sketch your ideal thermostat design in the space provided” on a sheet of paper. The experimenter explained to the participant that a quick sketch/drawing that illustrates the essential elements of the main interface would be sufficient.

This gave us the 48 sketches that constitute the data analyzed in this paper. To summarize before the fact, in terms of reflective feedback, the results that we got were very different than those which were obtained from the other instruments. In other words, had we not asked for the sketches, we would have unknowingly left valuable information on the table.

“Quick and Dirty” Analysis of All Sketches

We started looking at the sketches one at a time. It quickly became clear that it would be more effective to spread all 48 out on a large meeting room table. Immediately a

wealth of information was available to the viewer. Motifs and patterns started to emerge straight away – a welcome relief after having transcribed, coded and analyzed 24 one-hour video tapes! Rearranging and grouping sketches based on certain themes provided an increasingly deeper understanding of the collection.

We first grouped the sketches into the four experimental conditions from which they emanated. After this initial sort, we then arranged the sketches from each of the three single conditions based on their similarity to the associated prototype they had been presented with, from most to least similar. For the multiple design condition, we grouped the sketches based on the prototype they most highly represented, with the hybrids sitting in between the larger groups. All 48 sketches are shown in Figures 6-9, sorted according to these criteria.

Out of the 12 people who saw just the Linear condition, 6 used a pair of linear axes - only three of which bore any similarity to the original prototype. The Tabular design, on the other hand, was the one most closely replicated: 11 out of 12 participants in that condition used drop-down menus in their own sketches. Of these, 10 were very similar to the original design. Furthermore, five participants who never saw the Tabular design came up with up-down arrows or drop-down menus, and 2 participants who never saw the Circular design came up with the idea of a pie menu or a circular display in their own sketches - although no such element existed in the prototype they viewed - further suggesting the good fit of such controls for our system.

The sketches from the multiple design condition reinforced these observations. Of these, four subjects used a circular display and six used the idea of drop-down menus or up-down arrows (four of whom produced sketches that are very similar to the Tabular design). However only one made a sketch that was similar to the Linear prototype. Taken together, these data reflect the same trends derived from the questionnaires (Table 1). They reaffirm that the Linear design was not the best fit, and that users seemed to prefer elements of the Tabular or Circular designs.

They also suggest that the less the original design is reflected in the user sketches, the less successful it is. In other words, more highly rated designs are more likely to evoke sketches that contain elements of those designs than those with low ratings.

We then compared the sketches from the multiple design condition with those from the three single conditions. These sketches were richer in that there was more variation from any of the original designs. Although three sketches were relatively similar to the Circular, four similar to the Tabular and one similar to the Linear prototypes, only one sketch directly replicated one of these original designs (the Tabular prototype) with no further enhancement. The rest embellished one of the original designs, either by introducing new elements, or borrowing from the other designs seen. Among these there were two hybrid sketches

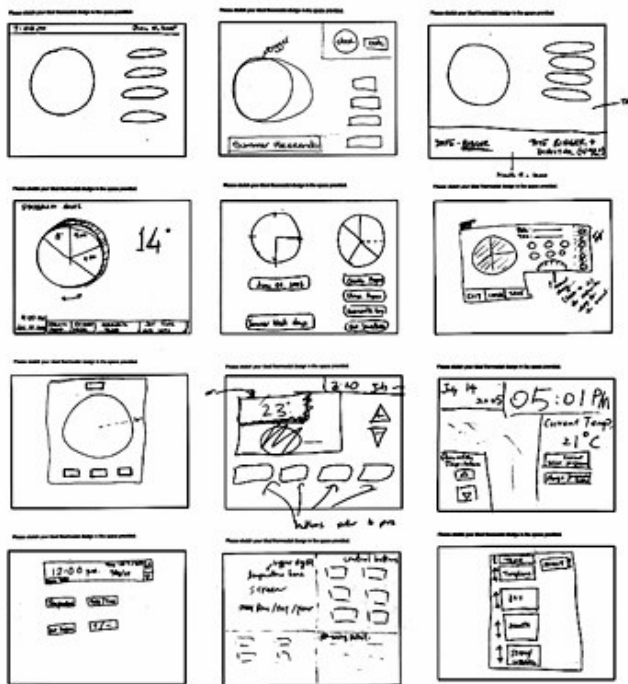


Figure 6: User Sketches from the Circular condition

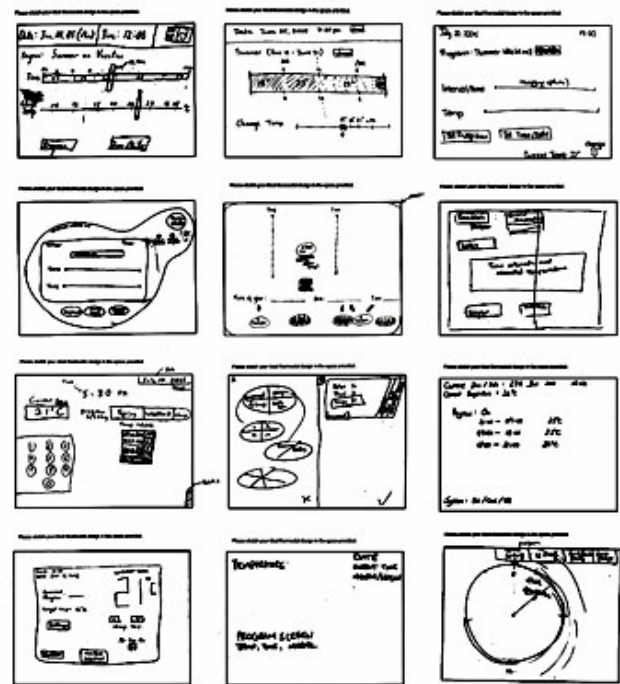


Figure 8: User Sketches from the Linear condition

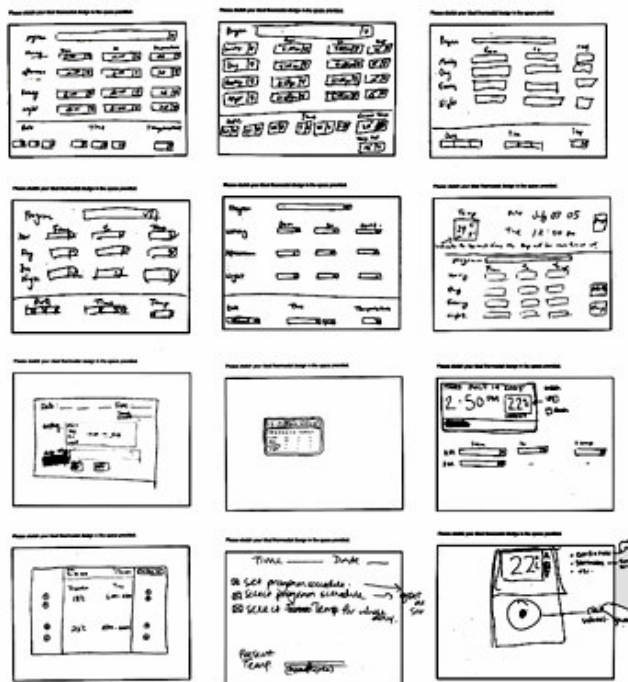


Figure 7: User Sketches from the Tabular condition

which bore no obvious similarity to any one prototype. Of course the participants in this condition were at an advantage as they were exposed to more than one prototype.

Further rearrangements and regrouping were done to explore other commonalities and differences; for example we were curious to find the number of digital *versus*

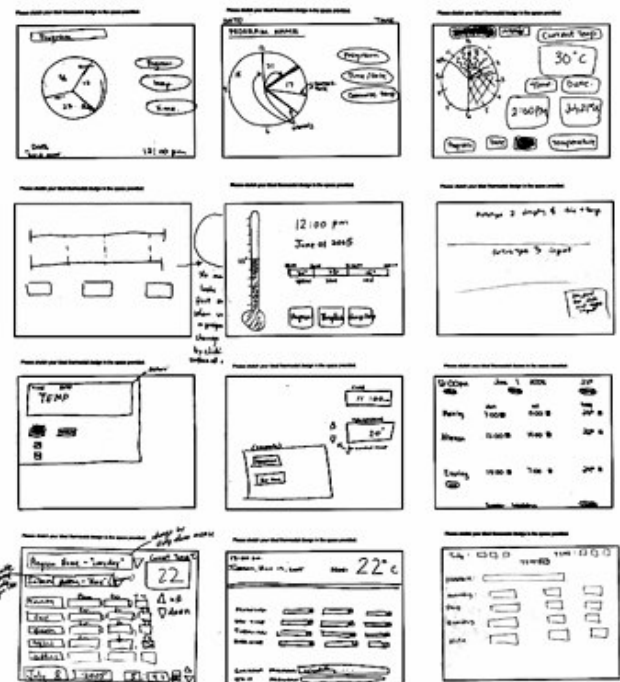


Figure 9: User Sketches from the multiple condition

analogue clocks, as well as the proportion of sketches in which the current temperature was directly specified on the main interface, as opposed to a secondary panel, as was the case in the Circular and Linear prototypes. It is important to note that the criteria according to which we sorted, clustered and categorized the drawings were not determined *a priori*. There were endless ways in which the sketches could have been rearranged or clustered,

Circular seen alone N = 12	Tabular seen alone N = 12	Linear seen alone N = 12	Multiple prototypes N = 12
8 used a circular display or dial (Out of those 7 were very close to original design)	11 used drop-downs or up-down arrows (Out of those 10 were very close to original design)	6 used a pair of linear axes (Out of those 3 were very close to original design)	4 used a circular display or dial (Out of those 3 were similar to the Circular design)
- 7 used a digital clock	- 10 used a digital clock	- 6 used a digital clock	6 used drop-downs or up-down arrows (Out of those 4 were similar to the Tabular design, and 1 directly replicated it)
- 2 used an analog clock	- None used analog	- None used analog	3 used a pair of linear axes (Out of those 1 was similar to the Linear design)
- 3 not clear	- 2 not clear	- 6 not clear	1 hybrid of Circular and Linear
- 6 showed current temp.	- 11 showed current temp.	- 9 showed current temp.	1 hybrid of Tabular and Linear
- 6 did not/was not clear	- 1 did not/was not clear	- 3 did not/was not clear	
<u>Original UI element examples</u>	<u>Original UI element examples</u>	<u>Original UI element examples</u>	<u>Original UI element examples</u>
-Buttons for On/off, Exit, Done, Cancel	- Buttons for On/Off, Default, Change, Auto-Setting	- Buttons for On/Off, Add-Remove-Intervals	- 8 used a digital clock
-Tabs	-Check Boxes	-Up-down arrows	- None used an analog clock
-Manual dial	-Manual Scale	-Flashing lights	- 4 not clear
-Visualization of steps for the pages in wizard	-Click-wheel	-Number pad	- 8 showed current temp.
-3D look for the dial	-Multiple screens, as opposed to one	-Pie menu	- 4 did not/was not clear
-Written instructions		-Circular display for temp.	
		-Wizard-based interface	
Acceptable	Acceptable	Least Favorable	Allows the exploration of hybrid designs

Table 2: Results of the “Quick and Dirty” Analysis of All 48 Sketches

thereby providing a rich source for insights. Our approach was fairly subjective – we mainly let the sketches themselves “tell us” things. This harkens back to the “seeing as” reading of sketches that we saw in Figure 1.

Finally all sketches were examined for introducing original UI elements. By original we mean elements that were not represented in the three prototype(s) viewed. Some of the most interesting additions were: tabs, click-wheel, manual scales or dials, On/Off buttons, flashing lights, and a thermometer display.

Table 2 summarizes the findings from our quick and dirty analysis of the 48 sketches.

One of the main advantages of this technique compared to those more commonly used in usability studies (e.g. think-aloud, questionnaires, interviews) is that performing the type of analysis described above is extremely quick given the affordances of the paper sketches. Unlike videotapes, it is possible to look at all of the sketches together.

Being able to recognize patterns at a glance, or answering unanticipated questions are unique benefits of sketches over numeric, textual, audio and video data. Experimenters can engage in a type of dialog with the sketches where questions are answered, answers are reflected upon, and new questions can be asked.

In-Depth Analysis of a Collection of Sketches

In order to better assess our sketch-based technique we compared the design ideas provided verbally in our first study with those illustrated in the sketches. In this we were limited to the earlier selection of 24 usability tests on which we had performed the in-depth video analysis. Thus, from here on we only consider the 24 sketches produced by those 24 subjects.

First, we were interested to find out how much time went into developing the sketches. No time limit was imposed or suggested to the participants. Participants spend an average of 3.9 minutes from the time that they were instructed to sketch, to the time they handed in their sketches; indeed a very short extension to the length of a full usability test.

Second, all ideas (redesign suggestions) previously generated throughout the think-aloud and interview (from the beginning of the experiment up to sketching), and those found in sketches were reviewed. What follows is an exploratory analysis based on counting ideas found in sketches vs. verbal protocols. The use of numbers here is to describe general trends and not to provide a statistical analysis, since we had no *a priori* hypothesis.

We counted a total of 84 ideas generated in the 24 experiments in question. Just over 21% of those (18 ideas) were applicable to the comparison with sketches. An idea was considered *applicable* if it had to do with the visual or functional aspects of the main interface, and could be sketched. Any ideas on use of colour, for example, were not applicable, as the user only had a single coloured pen to sketch. 89% of applicable ideas were found to be *repeated* directly in the sketches, and therefore would have been recognized even if the lengthy and expensive task of video analysis was omitted. Furthermore, 22% of applicable ideas were not only repeated but also *refined* (improved upon) during the sketching process. Only 2 ideas from the set of 18 (11%) could have been included in the sketches but were left out. The two ideas were: (1) adding a number-pad for data entry, and (2) adding an array of 12 buttons one for each of 1:00, 2:00...12:00. It can be argued that these 2 ideas would have been lost if we had not analyzed the recordings before the sketching experiment. However, as it will be shown in the next section, it could be the case that the participants willingly discarded these ideas after more reflection during the sketching process. In addition to previously mentioned ideas however, a large body of *new* ideas were formed during the sketching practice.

A *new* idea was identified as the addition of a new attribute to the HCCS (for example, adding a control that had not been previously encountered). In the multiple design condition a *combined* idea referred to borrowing an idea from one design and adding it to another. For example if the final sketch was similar to the Circular prototype but borrowed the idea of directly specifying the current temperature from the Tabular prototype, it was counted as a combined idea. This classification is consistent with previous work on ideation [15].

The new ideas generated by the users can be explained as discovering “new features and relations among elements that they have drawn, ones not intended in the original sketch” [11]. In order to find the magnitude of these newly discovered ideas we examined the 24 sketches and counted the number of new ideas included in the sketch that were never thought of, or at least never verbalized, before sketching. An impressive total of 38 new ideas were found. This means 68% of ideas that were interpreted from the sketches were new to sketches. Furthermore, six combined ideas were found in the sketches of the multiple condition that were not previously verbalized in the think-aloud or interview. These new and combined ideas would have been

lost if we had not allowed participants to sketch their ideal design.

Idea Type	Example
<i>Applicable</i>	“Having a temperature scale that is always visible for the Linear prototype” mentioned in the interview - (applicable because it can be sketched on the main interface)
<i>Repeated</i>	“Using digital instead of analog clock for the Circular prototype” - both mentioned in the interview and drawn in the sketch
<i>Refined</i>	“Having more specific features outside of the current options for the Tabular prototype” mentioned in interview without specifying what features, and then adding 2 such specific features in the sketch
<i>New</i>	Introducing a number pad to the Linear prototype - not previously mentioned in the interview
<i>Combined</i>	Borrowing the digital clock from the Tabular prototype and adding it to the Circular prototype - not previously mentioned in the interview

Table 3: Idea Categorization

These findings suggest that the simple act of sketching has enabled the participants to refine already generated ideas and discover new ones. Hardly any previously mentioned ideas were lost in the practice of sketching. The results were both complementary and reinforcing/supporting, relative to that obtained from the more expensive analysis of think-aloud and user interviews. They suggest that one needs to carefully consider when it is justified to employ more expensive techniques such as formal video analysis rather than lighter weight sketching techniques.

A Closer Look at Some Examples

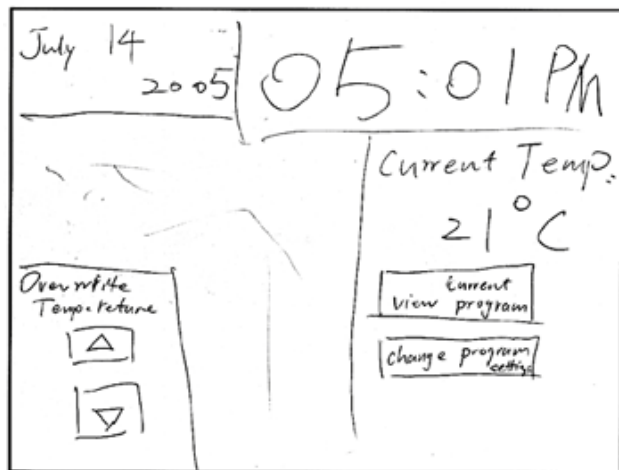
In this section we look more closely at a number of examples that illustrate the benefits of user sketching.

The Quiet Sketcher: Our first example comes from a participant (*P1*) who was only presented with the Circular design and rated it quite highly (9 on a scale of 1 to 10). He listed four likes and two dislikes for the design during the interview. However when asked “*If the prototype design was going to be used for the final system would you change anything?*” he replied “*No*”, instantly. The interviewer gave a second chance for communication (“*Any other suggestions?*”) and the answer was once again “*No*”.

When asked to sketch his ideal thermostat design, *P1* quietly stared at the blank paper for roughly half a minute, and then started his design from the top-right corner. In the next three minutes and 17 seconds, *P1* expanded the design to cover the whole page, without initiating any verbal communication. During sketching there were multiple

intervals in which he lifted the pen off the paper and reviewed the design, before returning to his sketching.

When turning in the final sketch, shown below, P1 explained:



P1: "That's about it. I have some space here. I guess the squares are too small" (pointing to the space in the middle and with a hand gesture showing that the four corner squares would join to cover the empty space.)

While P1 indicated during the interview he would not change anything about the design of the Circular prototype, his resulting sketch was fundamentally different from the original design and suggested a number of redesign proposals such as:

- (1) Removing the circular display representing the current program from the main page and instead adding a "View Current Program" button.
- (2) Providing a digital clock instead of an analog one, while significantly enlarging the representation of time.
- (3) Specifying the current temperature on the main display.
- (4) Enabling the manipulation of the current temperature directly on the main screen through the use of up and down arrows (both the concept and control are new to the Circular prototype).

The large white space in the middle indicates that P1 did not have a pre-defined design in mind when he started sketching, and rather arrived at this in the process of sketching.

The Passive Sketcher - with a Surprise Ending: Our second example is participant P2, who was only presented with the Tabular prototype. Her overall rating for this prototype was also 9 on a scale of 1 to 10. When asked in the interview to list her likes and dislikes of the design she listed four likes and two dislikes. When asked if she would change anything in the design for the final system, she simply said "Just the things that I found confusing" without elaborating on how she proposes to change those aspects. The interviewer had to use further probing and finally

verbalized a redesign suggestion on behalf of P2 based on her earlier comment of what was confusing:

Interviewer: "So you would activate it so to make all these [intervals] change at the same time... anything else?"

P2: "mmm. That's about it."

Interviewer: "OK, any other suggestions?"

P2: "No, I think that's it."

When asked to sketch her ideal design P2 was hesitant; "Oh my God, I don't even know what thermostats usually look like! I've only, like, used one Ever! I haven't drawn in a long time!", but began to sketch nonetheless.

P2: "I don't care as long as it has a big screen"

P2: "if it has [a big screen], it would...you could change this really easily."

P2: "and then what ever... this business" (pointing to the 4 intervals in the prototype indicating that it would remain the same in her design.)

Interviewer: "you can just quickly sketch it..."

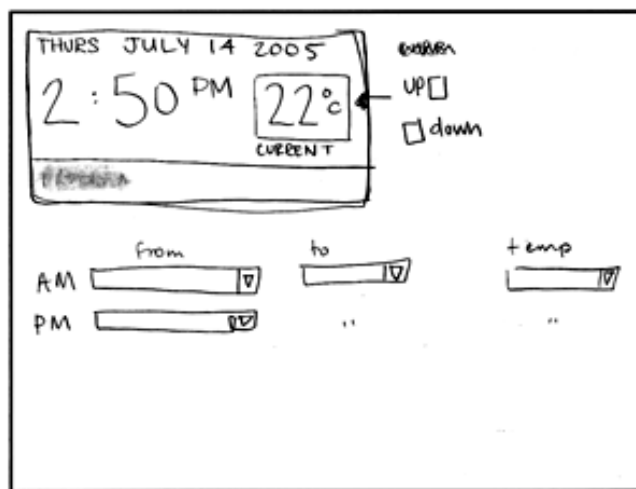
P2: "That's it. They are the same", providing further confirming that they will remain the same in her sketch.

However the final sketch (shown below) is surprisingly different from the Tabular prototype.

Although P2 was hesitant to sketch at first, and then repeatedly indicated that her sketch would simply copy the interval break-down ("this business") from the original design, she seemed to have discovered a totally new solution for representing the intervals.

Interviewer: "Oh, so instead of 4 [intervals] you have AM and PM?"

P2: "Yeah, yeah! At first I thought I would like that, but then I changed my mind"



This is an example of reflection on previous ideas and changing earlier reactions. Besides reducing the four daily intervals to two, this sketch suggests the importance of dedicating a large part of the screen to the current time and temperature as well as showing the day of the week (e.g. Thursday). A close look at the sketch reveals other signs of reflection and refinement (a scratched word above the Up button and what appears to be the residue of an erased element under the clock).

The Overly- Excited Sketcher: Our final example (subject P3) was presented with the Linear prototype only. Her overall rating for this prototype was 8 out of 10. Her attitude towards the design was overly positive with eight listed likes and one dislike. When asked what she would change about the design, P3 provided some suggestions but seemed to lose her train of thought easily during the interview.

P3: *"You know, when you go in the internet, when you touch, like, the day - today or something like that - then a calendar comes up and then you can just like touch the calendar rather than setting it? Maybe that's going to be easier, well definitely, it's easier!"*

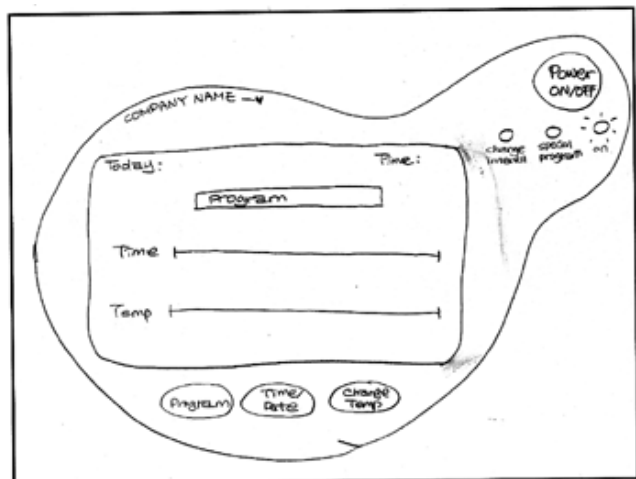
Then suddenly jumps to another topic:

P3: *"I don't know, I'm just so amazed by you guys actually program this! I mean I hated C, that's why!"*

And yet another:

P3: *"One second! I am just wondering, who are the target people? Like, just adults or their kids too?"*

When asked to sketch her ideal design, P3 immediately makes a radical change by turning the square screen into a circle! This is P3's sketch:



P3: *"mmm... do you need a square screen?"*

Interviewer: *"No, you could change that!"*

P3: *"I just didn't want to have the square one! ...Maybe I can put the power here... ah, you know this is cool! I get*

way too excited over this! ...the power and then there's like flashing (lights) for changing interval so you know like temperatures are changing so it's going to blink for a minute or something like that."

P3 was certainly excited to do the sketching exercise. Although during the interview she would flit from topic to topic, she was very systematic and organized when she turned to sketching. She sketched all details down to multiple flashing lights, even providing a space for the company logo! Her final remarks as she was handing out the sketch were: *"Thank you. I had fun!"*

The examples above illustrate a number of ways in which sketching helped the participants to better organize their thoughts, come up with new and unexpected design ideas, reflect on and refine previously stated ideas, and communicate their ideas to the experimenters.

DISCUSSION

This study illustrates the benefits of engaging users in a sketching activity as an extension to conventional usability testing. The act of sketching proved to facilitate reflection and discovery better than the other methods used. It has been known that "expressing ideas in a visual-spatial medium makes comprehension and inference easier than a more abstract medium such as language" [13]. Designers take advantage of this when they communicate their interface designs to users by the use of prototypes. However, participants of a UT test are typically expected to generate and communicate their own ideas using spoken or written language, which are (a) limited to describe concepts that are visual and spatial [13], and (b) do not facilitate the form of reading and interpretation ("seeing as" reasoning) represented in Figure 1.

We found that the sketching exercise resulted in receiving more reflective feedback such as design proposals, as opposed to reactive feedback such as praising or criticisms of the design. In our study we found that spreading all sketches on a table and looking at them as a whole was very effective. In traditional design practices displaying numerous sketches or designs on large foam/cork boards is very common. It is when sketches are seen in a larger group that patterns emerge and relationships are discovered. This method was also very quick, as most questions could be answered in a matter of minutes just by looking at the groups of sketches displayed.

Another major benefit of sketches for the researcher is that sketches allow for increasingly deeper interpretation and analysis. Questions such as "What percentage of users prefer digital or analog clocks?" or "How important is it for users to directly specify the current temperature, even if it can be found from the overall program?" were not anticipated prior to experimentation, hence were not incorporated into the questionnaire and interview; therefore finding these simple answers would have been difficult, if not impossible, if it were not for the sketches made by

participants. Like any other form of design, experimental design can benefit from reflection and iteration. It is not uncommon for researchers to start analyzing data only to find more interesting questions they could have asked! Often it is too expensive to redo experiments all together. One of the benefits of sketches is that they are not as limited as numeric or verbal forms of feedback, and allow for a form of dialog that enables one to think of new questions and find new answers.

All of the above findings suggest that the sketching exercise is a useful and effective addition to common usability testing practices. Furthermore, we showed that reviewing user sketches alone could have uncovered much of the essential findings of more commonly used UT techniques. More importantly it did so at a fraction of the time and money required to facilitate, record, and analyze the think aloud protocol, interview and questionnaire data.

CONCLUSIONS AND FUTURE WORK

In this paper we proposed user sketching as an extension to conventional usability testing techniques. We showed that sketching facilitated thinking, reflection and discovery for UT participants and resulted in receiving more reflective feedback. As such we recommend the use of this practice as a quick and inexpensive addition to other commonly used UT practices.

The high costs of UT limits the number of usability tests that one can undertake, and thus inhibits their use early in the design process, where one is rapidly exploring various alternatives and iterating through designs. We believe that the sketching exercise described has the potential to be developed into a light-weight form of usability testing, due to its relatively low cost in time demands for analysis. We hope that this study will be a catalyst for future studies on the practice of sketching in User Centered Design.

While this technique helps us get more out of conventional UT techniques, our research is directed at facilitating the early exploration of design alternatives with the involvement of users. As such, one of our longer term objectives is to explore such techniques further in ideation, especially when comparing multiple design alternatives.

ACKNOWLEDGMENTS

This research has been partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). This support is gratefully acknowledged. We would also like to thank Barry Po, Andrew Warr, Gilbert Cockton and Gayna Williams for their input and feedback.

REFERENCES

1. Catani, M.B. and Biers, D.W. (1998). Usability evaluation and prototype fidelity: Users and usability professionals, Proc. Human Factors and Ergonomics Society 42nd Annual Meeting, 1331-1336
2. Goldschmidt, G. (1991). The Dialectics of Sketching. Creativity Research Journal 4(2), 123-143.
3. Gould, J., Conti, J. & Hovanvecz, T. (1983). Composing letters with a simulated listening typewriter. Communications of the ACM (CACM), 26(4), 295-308.
4. Hornbæk, K., and Frøkjær, E. (2005). Comparing usability problems and redesign proposals as input to practical systems development, Proc. CHI 2005, ACM Press, 391-400.
5. Nielsen, J. (1993). Usability Engineering. Academic Press.
6. Preece, J., Rogers, Y. and Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction, John Wiley & Sons.
7. Schon, D.A. (1983). The Reflective Practitioner. New York: Basic Books.
8. Sefelin, R., Tscheligi, M., and Giller, V. (2003). Paper prototyping – what is it good for? A comparison of paper-and computer-based prototyping, Proc. CHI 2003, ACM Press, 778-779.
9. Smith, H., Fitzpatrick, G. and Rogers, Y. (2004). Eliciting reactive and reflective feedback for a social communication tool: a multi session approach. Proc. DIS, ACM, 39-48.
10. Snyder C. (2003). Paper Prototyping - The Fast and Easy Way to Design and Refine User Interfaces, Morgan Kaufmann .
11. Suwa and Tverskey, (2002). External representations contribute to the dynamic construction of ideas. In M. Hegarty, B. Meyer, and N. H. Narayanan (Eds.), Diagrams 2002. NY: Springer-Verlag, 341-343
12. Tohidi, M., Buxton, W., Baecker, R. & Sellen, A. (2006). Getting the Right Design and the Design Right: Testing Many is Better Than One. Proc. CHI 2006, ACM Press.
13. Tversky, B. (2002). What do Sketches say about Thinking? Proc. AAAI Spring Symposium on Sketch Understanding. 148-151.
14. Walker, M., Takayama, L. and Landay, J.A. (2002). High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes, Proc. Human Factors and Ergonomics Society 46th Annual Meeting, 661-665.
15. Warr, A. and O'Neill, E. (2005) Understanding Design as a Social Creative Process. Proc. 5th Conference on Creativity & Cognition, 118-127
16. Wiklund, M., Thurrott, C., and Dumas, J. (1992). Does the Fidelity of Software Prototypes Affect the Perception of Usability? Proc. Human Factors Society 36th Annual Meeting, 399-403.