GROUP PROJECT REPORT

Nyx

Assistant to sleep better

User Centered Design
University of Fribourg
December 2018

Hichem Belhocine
Donat Ajro
Timo Spring

Index

1. Project description

In the following report we present our user assistant application *Nyx* that helps people to sleep better and improve their sleeping habits. In the first part of this report, we briefly present the project and describe how we select the assistant topic and how if fulfills a user's need. Furthermore, we present the persona and scenario. In the second part of the report, we state our design decision and elaborate on our decision making. In chapter three, we are stating our evaluation plan and hypothesis. In the last chapter, we present our evaluation results and interpretation and conclude with a project retrospection.

1.1 Needs and requirements

We propose three initial project ideas - eating healthier, sleeping better and exercising more. To decide on the project, we conduct a survey with 13 questions impacting most topics. The survey is filled out by 31 randomly selected people. In the survey, we also directly ask the users, what application type would be preferred resulting in a clear picture as seen in Fig. 1.

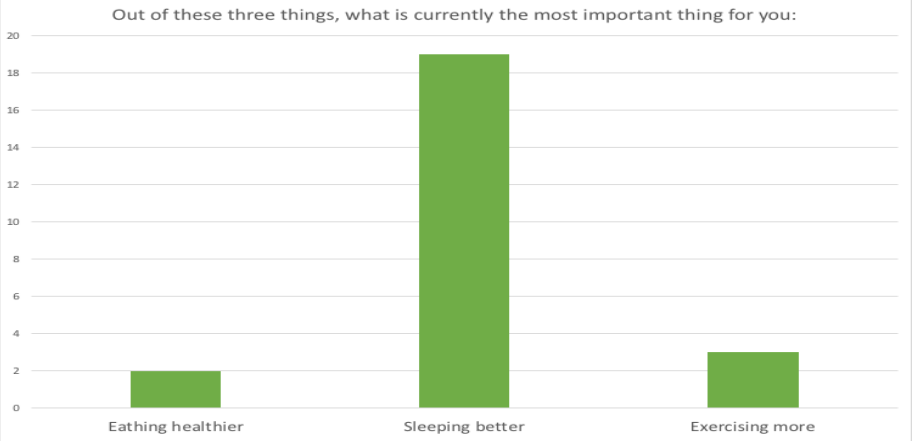


Fig 1. - preferences in application assistant type

We conclude that there is a clear need for an application that supports the user to sleep better. Therefore, we choose the sleep assistant as our project.
During a brainstorm phase we define the following application requirements. The application should remind the user when to go to bed depending on a selected time slot. To fall asleep faster, the app can play some relaxing sounds like ocean waves that help the user relax. The user can also set the alarm clock over our application and set wake up tones and different modes to be woken up, for example a soft or brutal wake up call. Furthermore, we provide additional features such as snoring prevention and nightmare detection based on the movement and noises the sleeper makes. On a general overview screen, we display sleep related articles that might interest the user. Furthermore, we provide an overview of different sleep statistics such as sleep duration, movement and noises.

1.2 Persona

Our persona is named Laura Favre. She is blond, 22 years old and an undergraduate student in media and communications (4th semester) at the University of Fribourg.
She lives in Fribourg in a dorm with 2 other people - Tom, the sports student and Eve, the physics student both 24 years old. Laura usually does not find time to cook, so Tom cooks for them on a daily basis, since he is a sports student and does not have to do much homework in the evening. Unfortunately, Laura is snacking during the day and in the evening, while Tom is preparing a dinner for them. The dinner is usually very healthy since Tom eats only healthy stuff, but her snacking habits are terribly unhealthy (lots of chips and chocolate and energy drinks). Laura used to play volleyball but now she does not have enough time to play it on a frequent basis.
Laura does not get a lot of sleep, since she spends most of her nights catching up with homeworks. This is one of the reasons that she has troubles to fall asleep, because when she goes to bed after her learning sessions, her brain is still overloaded with homework. Furthermore, she uses her phone extensively to check the social media activities of Tom right before she goes to bed. During the night, Laura wakes-up frequently to go to the bathroom. In the morning, she has problems getting up and is a frequent user of the snooze button on her alarm clock. Oversleeping therefore happens often. To feel less tired, she drinks lots of energy drinks in the morning to get her engine started.
However, she feels like her current lifestyle is putting a lot of pressure on her body and mind. Furthermore, Tom disapproves with her lifestyle which frustrates her even more, since she would like to impress him at all times. She would like to exercise twice per week so she gets in form. She drinks only one litre of water in average per day and at least four energy drinks. On the weekends she usually eats out with friends at McDonalds and then goes on to drink in bars or discos until early morning hours. She uses her iPhone 6 and a MacBook Air mainly for the university work and her social media activities. However, her MacBook Air has a poor battery life and she is therefore quite dependent on her iPhone. Since she wants to be reachable for Tom all the time, she keeps the phone in her pocket at all times, it's complicated.

1.3 Scenario

Laura gets home from university at 19:00, eats dinner and starts working on the homework until 22:30. Our app then notifies her that she has to go to sleep in 30 min and should turn off her digital devices now, so her brain will not stay awake in the bed. It also suggests that she sets up the alarm clock for the next morning. She prepares to go to bed, brushes her teeth and puts on her pyjamas. Because of the old habits she tries to pick up the phone to do the usual social media tracking of Tom, but fortunately for her, our app has that covered and blocks her from using the phone for such activities before sleep. Laura goes to bed. As she is trying to fall asleep the app plays the ocean waves relaxing sounds. Since she also has chosen a program from our dream-coach, the app starts playing the chosen program

after a few minutes into the ocean waves. “You are great, you can achieve everything!” is played back softly in the background only disrupted by the occasional ocean wave sounds. During the night, the app monitors the sound levels and can detect that the Laura is experiencing a nightmare, so the app starts ringing softly to wake her up enough to leave the nightmare. Since Laura chose the “Best” wake-up program from our app, the app can detect the soft sleeping phase around 7:53 A.M and decides to wake her up by playing the soft alarm sound and simulating the sunrise with the flashlight. She wakes up, picks up the phone and the app provides the statistics of the last night’s sleep.

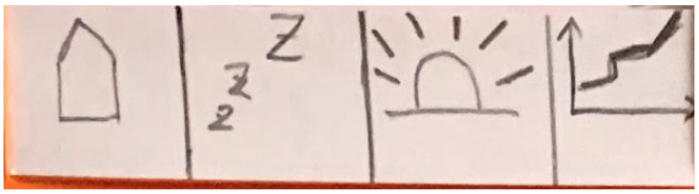
2. Design Process & Decision Making

After the initial survey and requirements specification and user needs definition, we start brainstorming on our application design. We started breaking down the possible activities based on our requirements to possible design solutions such as drop downs. A first key decision, is what the main features are to be displayed prominently. Based on the user survey, we choose the main features to be the time picker and reminder for the bedtime, the alarm clock for the next morning, and the sleep statistics. We consider putting all the features into a single screen - thus forcing the user to scroll to discover all features. This might reduce the general discoverability a bit but could positively impact the efficiency, since it would potentially generate less clicks. However, since we also want to display more advanced features, such a single page approach can quickly become crowded and complex, negatively impacting the simplicity and overview. We decide to try a different approach and categorize the main features including their advanced options in groups of logically similar topics. This results in the following groups:

- Home: including a sleep score (summary of different sleeping metrics), the sleeping period, and different sleep related articles and references to our main features.
- Bedtime: includes a time picker for the bedtime and some notification settings for the bedtime reminder. More settings include the dream coach, snore detection toggle and nightmare detection toggle.
- Wake-up: includes setting the alarm clock for the next morning and choosing between different wake up modes, wake up ringtones and whether to simulate the sun rise with the flashlight of the smartphone.
- Statistics: display all sleep related metrics such as sleep duration, noise level during the night etc.

With these groups in mind we decide to create a menu bar and place it at the bottom of the screen (Fig 2.), since the reachability is best there. Each item in the menu bar takes you into a different tab where the grouped feature is presented.

Fig 2. -Menu bar



The starting page in our case would be the home tab. The idea to include a home screen is to have one place where all information is collected, like it would have been in the single page approach. We also include some articles and studies on sleep matter, for example, an article on a study indicating that heavy smartphone usage before the bedtime leads to problems falling asleep.

Next design decision affects the order of the menu bar items. We compare different native sleep tracking applications and native alarm clock applications to profit from possible mental models. We put the home screen at the first position starting from left to right. In addition, we simulate the user's journey through the day in the menu bar order. As the day passes the

user would want to set the time on when to go to bed, thus comes the bedtime icon on the menu bar and the screen as the second order. After that the user can set the time on when to wake-up hence the wake-up as the third order and as the last thing after the user wakes up can start checking the statistics on how was his or hers sleep.
For these design decisions we make heavy use of paper prototypes as can be seen in Fig 3. They facilitate easy testing with users and help visualizing the different approaches.

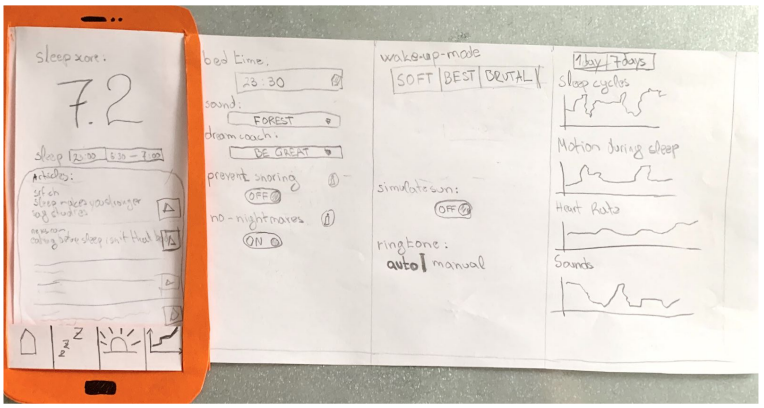


Fig 3. -Pape prototype

3. Evaluation plan

3.1 Topic Selection

An important feature in the application *Nyx* is the ability to set the bedtime and the alarm clock for the next morning. We propose a design to perform both actions on a single interface. Thus, the user's awareness of the upcoming sleep duration already when setting the times in the app. This might further motivate to have an early night, because the sleep duration becomes more tangible. Furthermore, a warning could be shown if the sleep duration is below a certain threshold defined over the user's age.
In addition, having both time pickers - for bedtime and the alarm clock - on the same page increases the efficiency and usability, as the user does not have to lock for both settings in different locations.
Native alarm clocks known from iOS or Android are usually displayed in the form of a time picker wheel. There is one wheel for the hour and one wheel for the time selection. By scrolling on the hour and minute respectively, the selected time can be changed. However, in the native OS we also find other approaches for time pickers. iOS also offers a sleep analysis feature, where you set the bedtime and the alarm clock for the next morning. However, this feature is displayed using a traditional watch face with coloring the range between the bedtime and the wake up time as the sleep phase. The times can be changed by selecting the beginning or the end of the sleep phase and dragging and dropping it to a new time. Hence, this approach is fundamentally different from the time picker wheel. In certain Android versions, one can find similar time pickers for the alarm clock. For an evaluation type we use the controlled experiment.

3.2 Hypothesis

Based on these two fundamentally different approaches to pick a time, we state the following null hypothesis.

> ***Hypothesis***
> *There is no difference in user performance (time and error rate) when picking the bedtime and setting the alarm clock from a time picker wheel interface or from a watch face interface, regardless of the subject's previous expertise in using a smartphone or using the different time picker types.*

The watch face UI feels more natural since it corresponds better to our mental model of a traditional watch and time. However, due to the heavy use of time picker wheels in native OS, the mental model might have changed. We are interested in knowing whether there is a significant difference in the performance when comparing these two interfaces. Our hypothesis states that there is no difference in the performance. To measure the performance, we enclose the time required to set the bedtime and the alarm clock and the error rate. To achieve unbiased results, we pick subjects randomly and not based on their

knowledge and expertise in the use with smartphones. Furthermore, we make sure to also include people with less expertise in using a smartphone.

3.3 Variables Definition

3.3.1 Independent Variable

Our hypothesis includes the independent variable that we will manipulate during the experiment independently from the subject's behavior. The independent variable being the time picker type - so, the time picker wheel or the watch interface. For that purpose, we prepare two high fidelity prototypes of our interfaces - one for the time wheel and one for the watch interface. This allows us to easily test both interfaces with the users. We try to simplify these interfaces as much as possible, so that they only contain the core functionality of setting the time - hence no decorations in the design are used. This ensures that the user is not biased because of different styling techniques.

3.3.2 Dependent Variable

We state several dependent variables that we want to measure during the experiment. To ensure exclusive dependency on our independent variable, we state the following performance factors:

- time to set the correct bedtime and wake up time (alarm clock)
- number of clicks (on touchscreen) required to set the times
- number of errors (wrong clicks)

We also include a satisfaction score in our experiment. Hence, the interface with the poorer performance factors could still be prefered by the users in terms of satisfaction score. One reason for such a situation could be that the interface with the better score simply feels more natural or that the design is preferred over the other. In addition, the satisfaction score provides a mean to the user to rate the interface independently of its performance.

3.4 Subject Selection & Assignment

For our experiment we chose the within-group design. So all subjects are assigned to all cases. It helps us isolating the effect of the individual differences and allows the users to better compare the two approaches with the satisfaction score.

As our subjects, we select select six people from mainly two different age groups, the younger and a bit older participants i.e. three participants between 20-29 and two participants between 50-59 and one participant younger than 16. For the gender, we have four men and two women. We form three groups with two participants for each group. To reduce the subject variability, we assign the two participants with the age group 50-59 in the same group. Therefore, the group itself becomes more homogenous. This selection and grouping of users helps to eliminate any biases that might stem from subject variability.

Using the within-group might have an impact on the fatigue level of our participants, since they have to perform the task twice - once for each interface type. However, since our task of setting the time is quite simple and should not take too long, this risk can be accepted.
At the end of each experiment session, the users have to answer some general questions about the sleeping habits and their personal relation to technology and their demographics.

3.5 Controlling Bias

For each group, we use a different interviewer. Hence, we are reducing the bias stemming from the perception of the interviewer. Furthermore, we prepare a script with the task to be read to the user and review it before the experiment. Therefore, all the interviewers give the exact same instructions - word by word. Questions by the users are not answered during the experiment to prevent the different responses from the different interviewers to effect the results. The metrics are all captured in the same format as described by the task description [Task Script].
For each group the participants are using the two versions of our time picker in a different order. So one person first checks the time wheel and then the watch face and the other person first the watch face and then the time wheel. Thus, reducing the learning effect. After one day, the experiment is repeated with the same participants, but in the opposite order.

3.6 Statistical Analysis

Since we are using the within-group design, we utilize the two-tailed paired-samples t-test to evaluate our results on a 0.5 significance level. This test is suitable, because we state one independent variable - the time picker type - with two conditions - watch face or wheel. The statistical analysis is conducted for each dependent variable first across all participants but we also calculate it separately for each group. Furthermore, we use bar charts to visualize our data to make out outliers or anomalies. And we add the mean calculation to these charts.

This chapter provided an overview of our experiment plan. In the upcoming chapter we will present the results of the the experiment.

4. Results

The following chapter presents the results of our evaluation. First we summarize them for each dependent variable. In the second part we perform a paired-samples t-test to check for significant differences between the watch interface and the wheel interface.

4.1 Evaluation Results

Each person tested both versions of our interface on two different days in different order. For the evaluation, we average the results for the same version from the the two days.
Fig. 4 presents the results for the completion time for both interfaces. The completion time in that context, is the total amount of seconds required to fulfill the task of setting the bedtime and the wake up time.

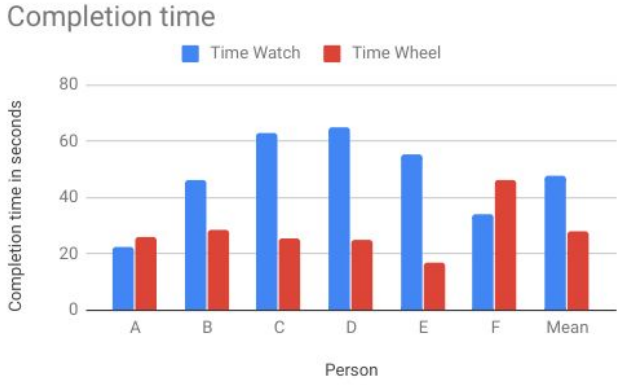


Fig. 4 - Completion time for both versions in seconds

We observe that the time to complete the task of setting the bedtime and wake up time for the watch interface is much higher than for the wheel design. In fact, the mean of the watch design is 47,7 seconds and for the wheel design is 28,1 seconds, so nearly half.
Looking at the number of clicks required to fulfill the task in Fig. 5, we observe a similar pattern. Thereby, each click on the touch screen was counted. For the wheel interface, scrolling on the wheel, was also taken into account as clicks.

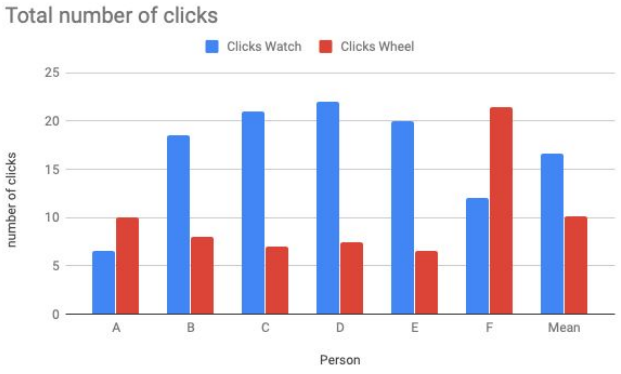


Fig. 5 - Total number of clicks required to fulfill the task

We state the number of clicks seem to correlate with the completion time. So, the more clicks the user has to do, the more time it requires to complete the task. The mean number of clicks for the watch design is 16.7 clicks and for the wheel design is 10.1.
Fig. 6 presents the error rate, so the number of wrong clicks in relation to the total number of clicks. For each version there is a path of clicks required to finish the task e.g. select save. A wrong click is considered, when the user leaves this path of clicks and clicks somewhere he is not supposed to.

Fig. 6 - total error rate in percentage for each version.

We observe in Fig. 6, that the error rate of the watch design is higher than for the wheel design, confirming the assumptions from the previous results. The mean error rate for the watch design is 52.6% and for the wheel design 42.3%. We note that the higher error rate might also explain the higher completion time for the watch interface and that this design in general is more error-prone.
To also take into account the user's perceived satisfaction with the two interfaces, we include the satisfaction score in our evaluation. The satisfaction score is rated on a scale of 1 to 5 with 1 being very unsatisfied with the interface and 5 being very satisfied with the interface.

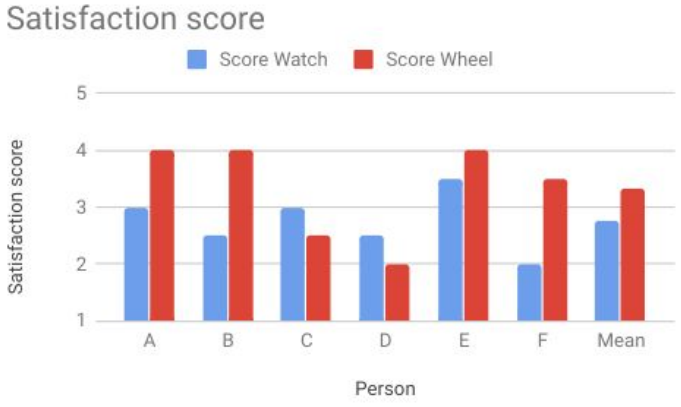


Fig. 7 -Satisfaction score on a scale of 1 to 5.

We note that the satisfaction score for the wheel interface is higher in most cases. One reason for this might be that the interface showed to be less error-prone and took less time to complete. However, it is interesting to observe that person D rated the watch interface higher, even though it took longer to complete and caused more errors. As a comment,

several users mention that the watch interface looks better in terms of styling and design, but is just more difficult to use and less intuitive. We also observe that several users were trying to drag & drop the time selectors in the watch interface instead of clicking on the time to select it. The mean satisfaction score for the watch interface is 2.75 out of 5 and 3.3 for the wheel interface.

4.2 Statistical Analysis

We perform the two-tailed paired-samples t-test for our dependent variables using the built in excel feature. This allows us to compare the results on a alpha-significance level of 5%. The null hypothesis is that there is no difference in the results of the watch interface and the wheel interface. Our two-tailed critical value is 2.57. If the t-value exceeds this critical threshold, the null hypothesis can be rejected.
For the completion time the t-value is 2.07. Therefore, we can not reject our null hypothesis that there is no difference between the completion time of the watch interface and the completion time of the wheel interface. The t-value for the total number of clicks is 1.54. Hence, we can also not reject the null hypothesis here. For the error rate, the t-value is 1.73. So again, the null hypothesis holds. For the satisfaction score the t-value is 1.55, hence there is also no difference there.
We also conduct the statistical analysis for each group separately. The individual results confirm the results in all but one case. For the group with the participants C and D, the difference in the completion time between the watch face and wheel face is significant (t-value 27.3 and critical value of 12.7).
So, over all our dependent values, there is no significant difference in the results with a probability of 95%, when looking at the results across all participants. The statistical analysis is available online.

4.3 Interpretation

Even though we observe some differences in the results pointing to the assumption that the wheel interface is performing better than the watch interface for all dependent variables, our statistical analysis does not confirm that. It shows that there is no significant difference between the results of the watch interface and the results of the wheel interface, except for one group. There we can observe that it takes significantly longer to complete the watch interface task. We state that even thought the watch interface was perceived more positively by the users in terms of look and feel, it is more complicated to use.

4.4 Subject Interviews

After the experiment we interviewed our participants with a focus on their sleeping habits and preferences in utilizing a smartphone application that helps them to sleep better.
Among other questions, we ask them how comfortable they feel using the smartphone close to the bed to track their sleep. Even though, all participants want to use a smartphone to use such an application, the majority feels uncomfortable using such an application.

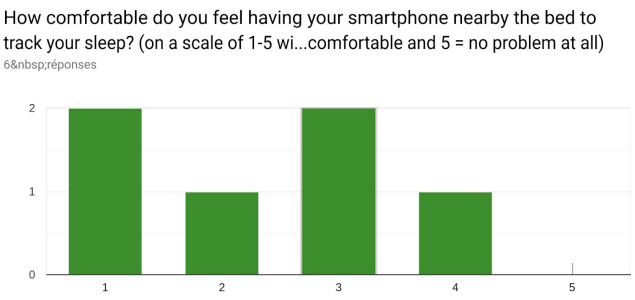


Fig 8 - comfortability level on scale of 1-5 in using smartphone to track sleep

At least 56% of our participants show issues in falling asleep and about 66% wake up in the night mostly to go and pee. However, the majority of the participants rates their sleep quality as good with an average sleep duration of about 6-7 hours in 66% of the cases.
When asked what the application should do, about 83% of the participants responded that they would like to have an assistance to help them feeling less tired in the morning. All the results can be observed here.

5. Project Retrospective

For the prototype of the watch interface we observed that our test participants nearly always tried to use drag and drop to change the time selected rather than just clicking on the prefered time. For a future version, we could use the drag and drop action, which should drastically decrease the amount of time to complete the task and also reduce the number of errors.

The first questionnaire gave us some interesting results that made us shift from one idea to another. The paper prototype proved to be very valuable in verifying key design decisions such as the number and structure of the menu tabs. And served as a reference onto coming up with new innovative features.

Annex

A - Task Script

Preparation:

- prepare stopwatch to stop the time
- open prototype on mobile phone:
- use the following form to count the dependent variables:
total time to fulfill task [in seconds]:
total number of clicks:
total number of errors:
satisfaction score:

Note:
You have to do the test with 2 people. Both persons check both versions of the time picker during the test but in a different order. After one day, they repeat the test but with the opposite order. That means:

Person 1: First test version 1, then version 2 on the same day.
After one day repeat the test, but first version 2, then version 1.
Person 2: First test version 2, then version 1 on the same day.
After one day repeat the test, but first version 1, then version 2.

Experiment process:

- Read task description to user
- stop the time and count the number of clicks and errors the user makes
- let user perform task
- stop time again and sum up the errors
- ask the user the interview questions
- wait one day and then re-do the same test with the same user (without the interview questions)

Task Description (to read to the user):

I will now read you a task that you have to perform using a special application. Please comment your thoughts during the test (think aloud) and elaborate on any unclear points. However, I will not answer any questions during the test.

Your task is as follows:
Please set up a timer, so that you are notified at 23:00 to go to bed and at 6:00 to get up in the morning. Please let me know when you feel like you have accomplished the task.

STOP TIME and select the correct version in the prototype. Hand the phone to the user and count all clicks the user makes (touch on display) and all errors the user makes (for example clicks on wrong number, clicks on inactive fields etc.)

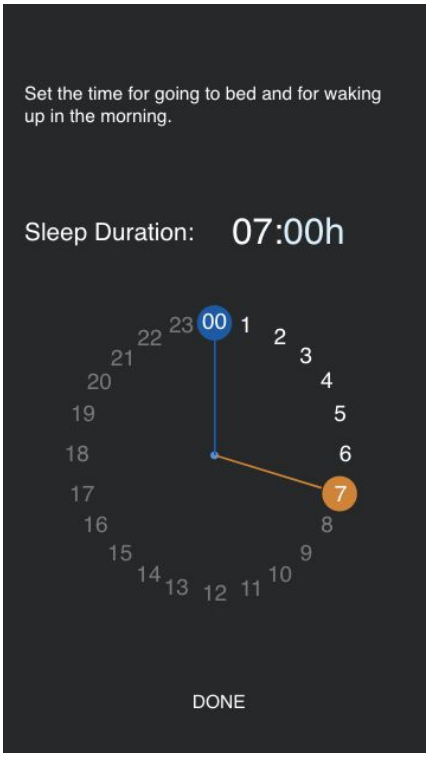
-> User you can start now!
STOP TIME when the user has finished.

Repeat
Redo the same task, but with the other version of the time picker.

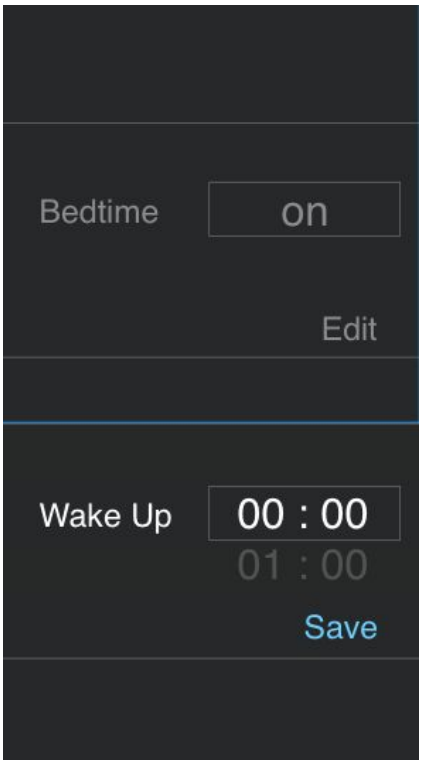
Interview:
After the user has tested both versions, ask him the interview questions.

B - Prototypes

Experiment Prototype: Can be found online.



Watch Interface



Wheel interface

High Fidelity Prototype: