Auroscope

Assignment 5: Lofi Testing

CS 147 Winter 2021

Pramod Kotipalli, Katie Pieschala, Nathan Sariowan

Introduction

Value Proposition

Capture your inspiration.

Mission Statement

We want to empower growing artists and musicians to feel more creative by making it easy for them to catalogue inspiration from their day-to-day experiences.

Problem/Solution Overview

"Writer's block" isn't just for writers: almost all people engaged in creative work feel stuck at times, as if the faucet of creativity just won't open up. We spoke to artists and musicians to discover a need for easily capturing digital media in all its forms (video, audio, drawings, etc.) and making that content useful for triggering creative processes.

We are developing a mobile application featuring a gesture-based interface for capturing media from many different phone sensors in one place. The captured multimedia is automatically synthesized into creativity-sparking formats like mood boards or vision boards. These collections or multimedia content can then be shared with friends who can collaborate with you or draw their own inspiration from others' content.

Sketches

Concept Sketches

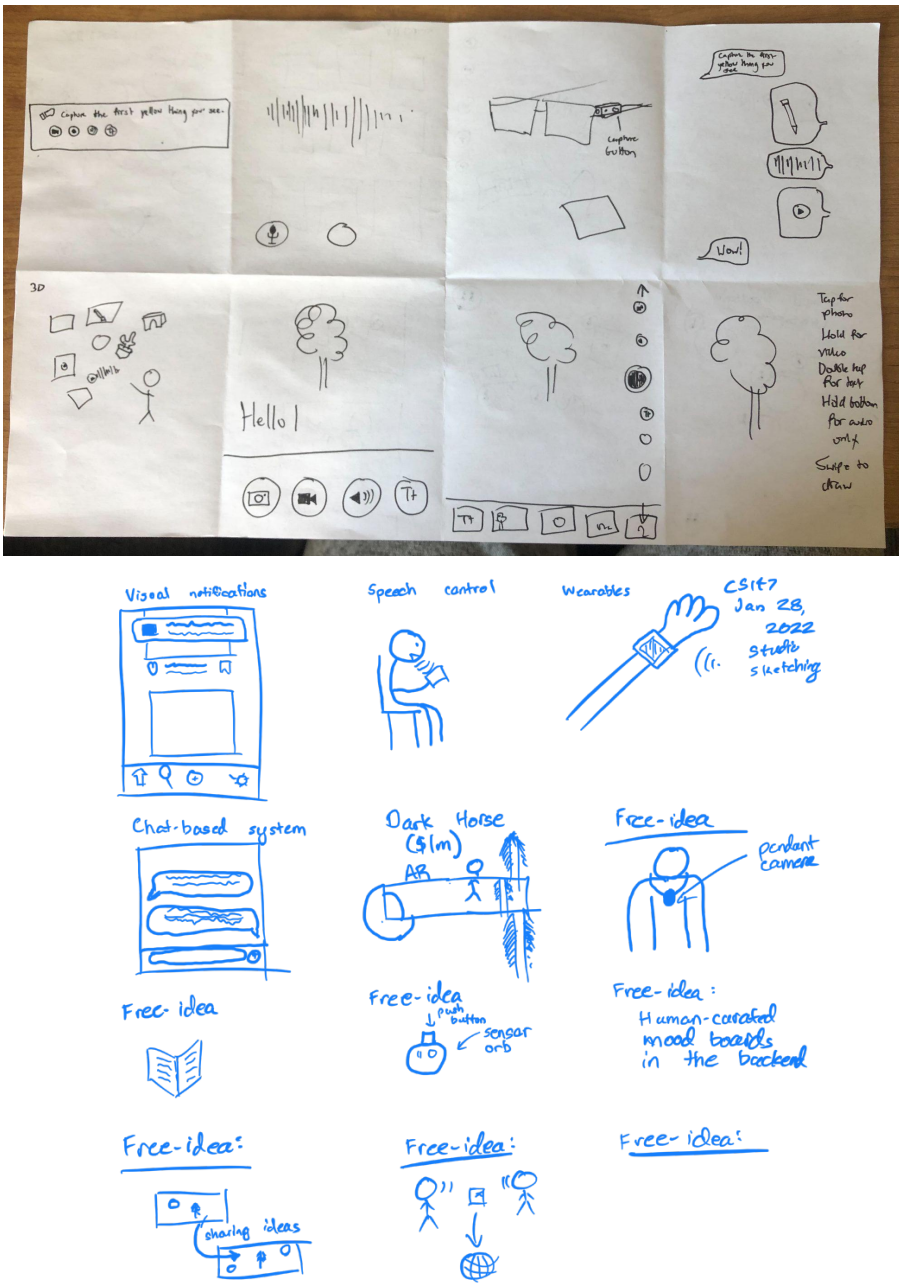


Figure 1: A series of crazy-8 brainstorms gave us several diverse realization ideas to go off of.

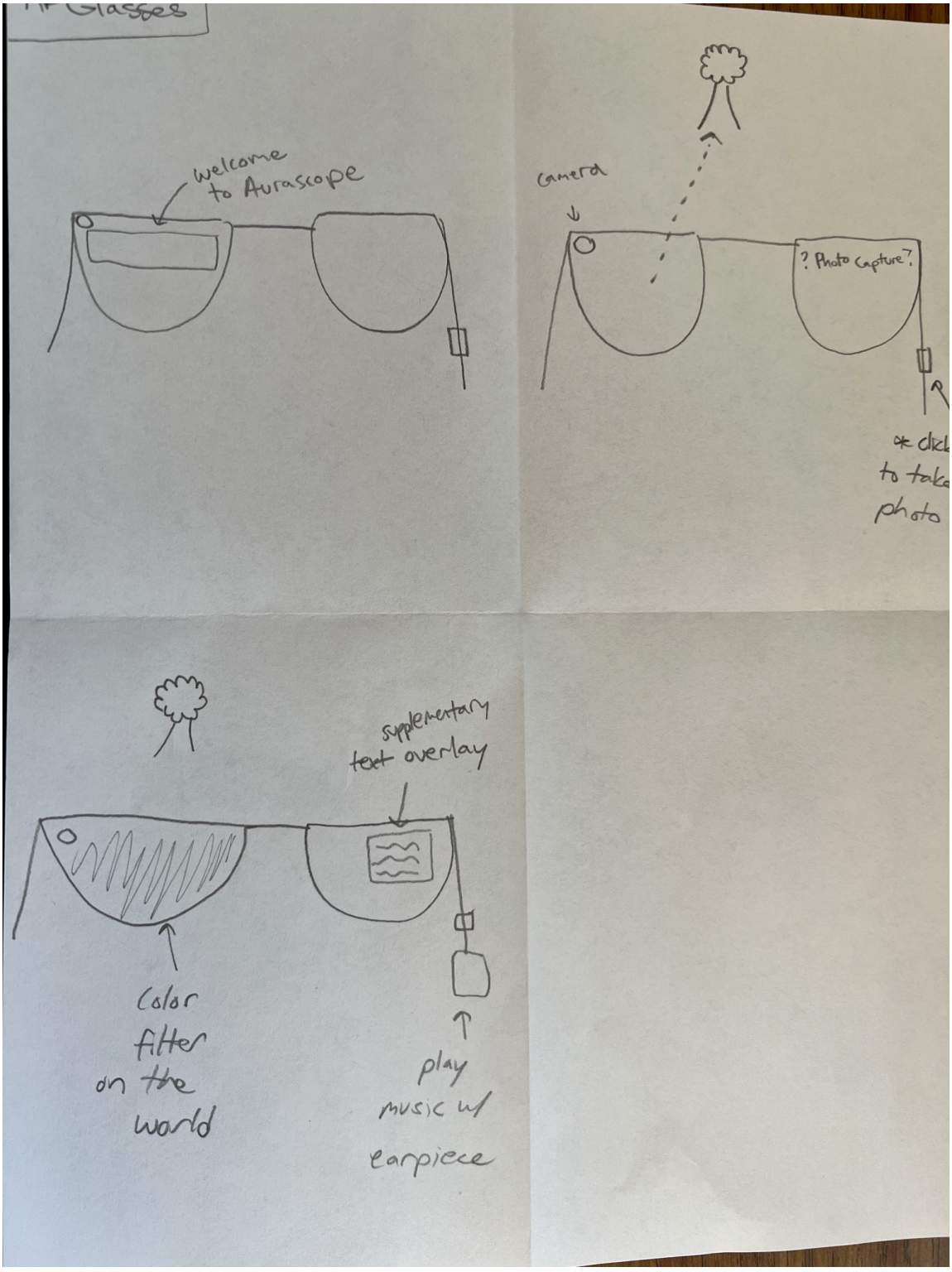


Figure 2: An idea for an AR glasses realization of Aurascope well embodied the mission of quick, life-integrated capture.

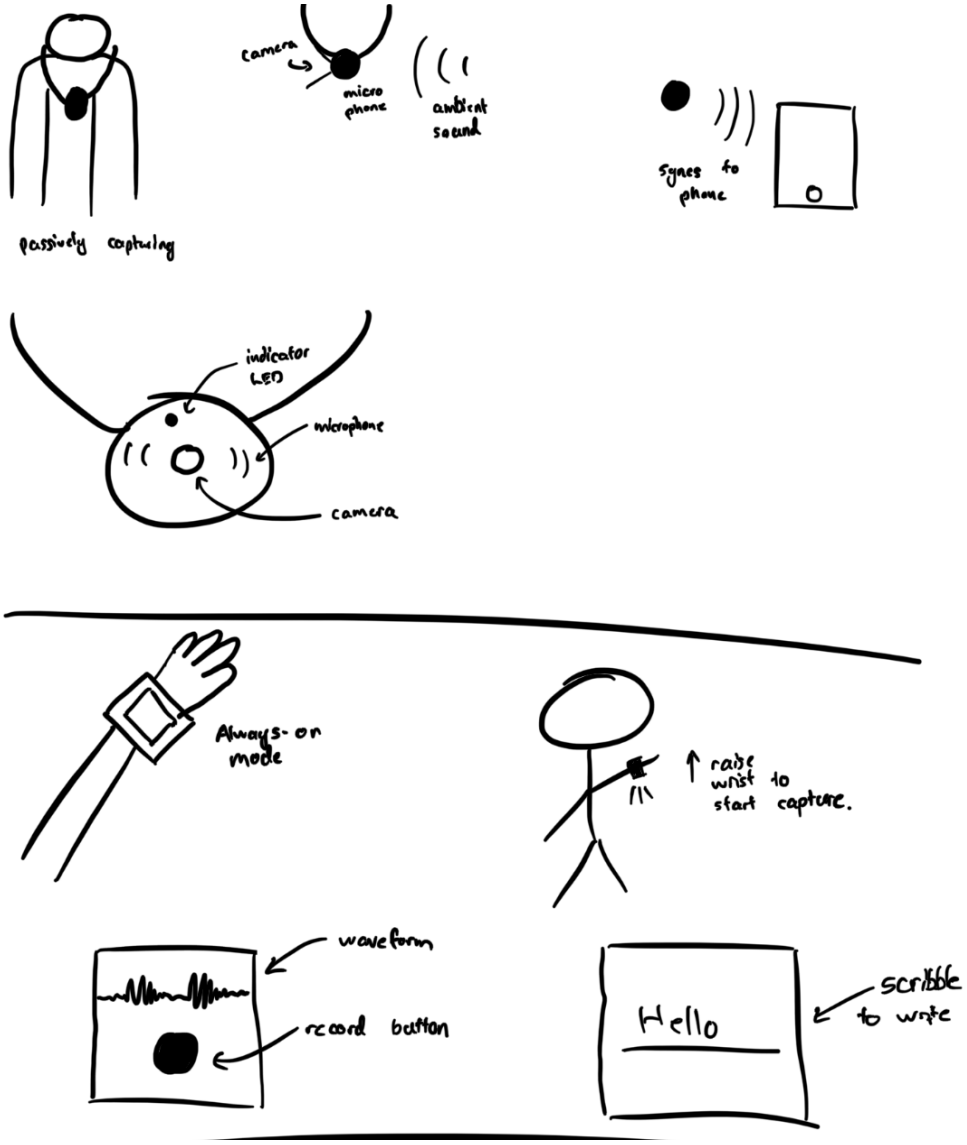


Figure 3: Two design ideas: wearable necklace and smartwatch concepts that automated the capturing process.

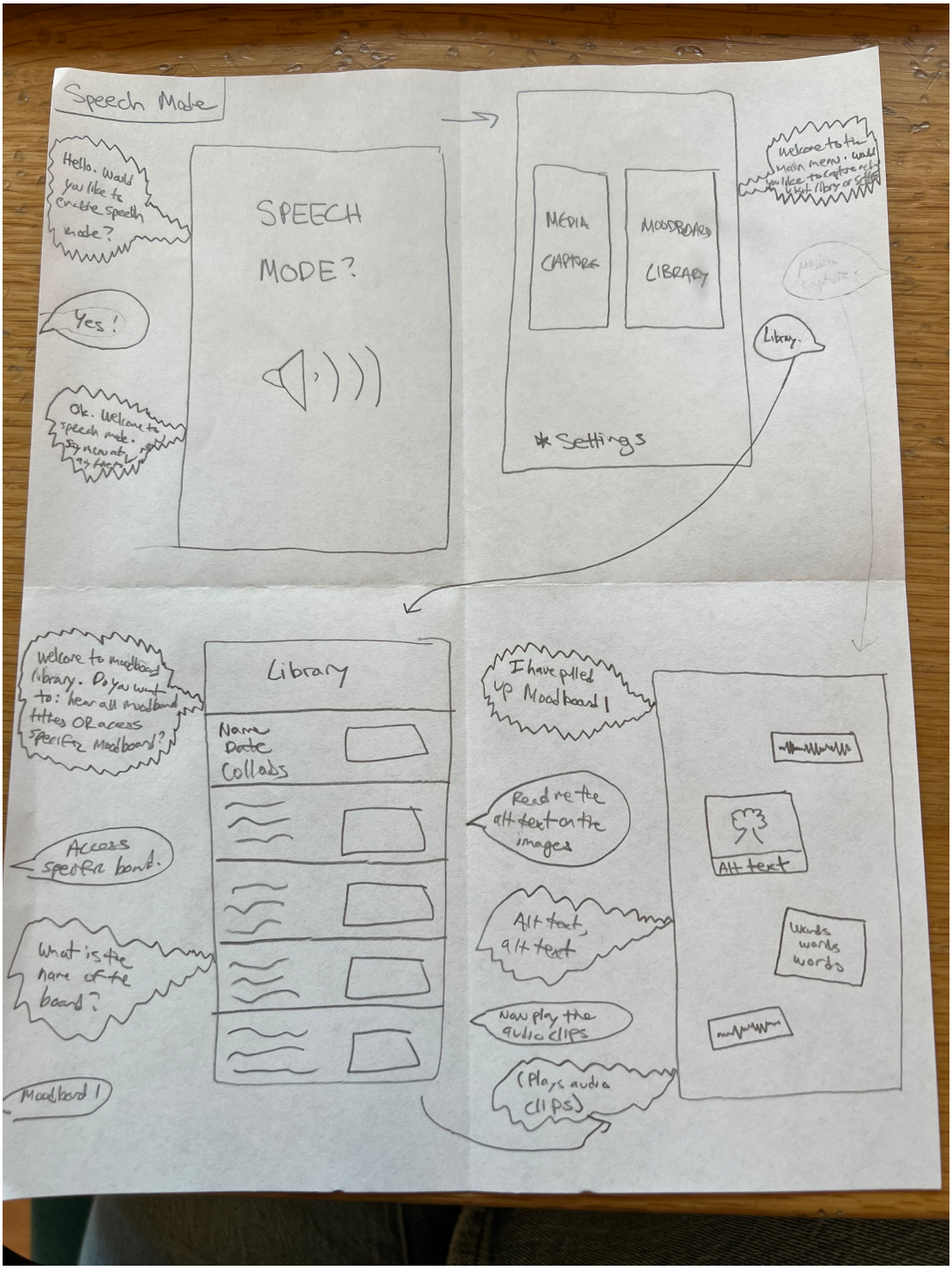


Figure 4: In designing with accessibility in mind, one idea we came up with was a speech mode that made the interactivity and automatic generation feel more natural under the spirit of a virtual assistant.

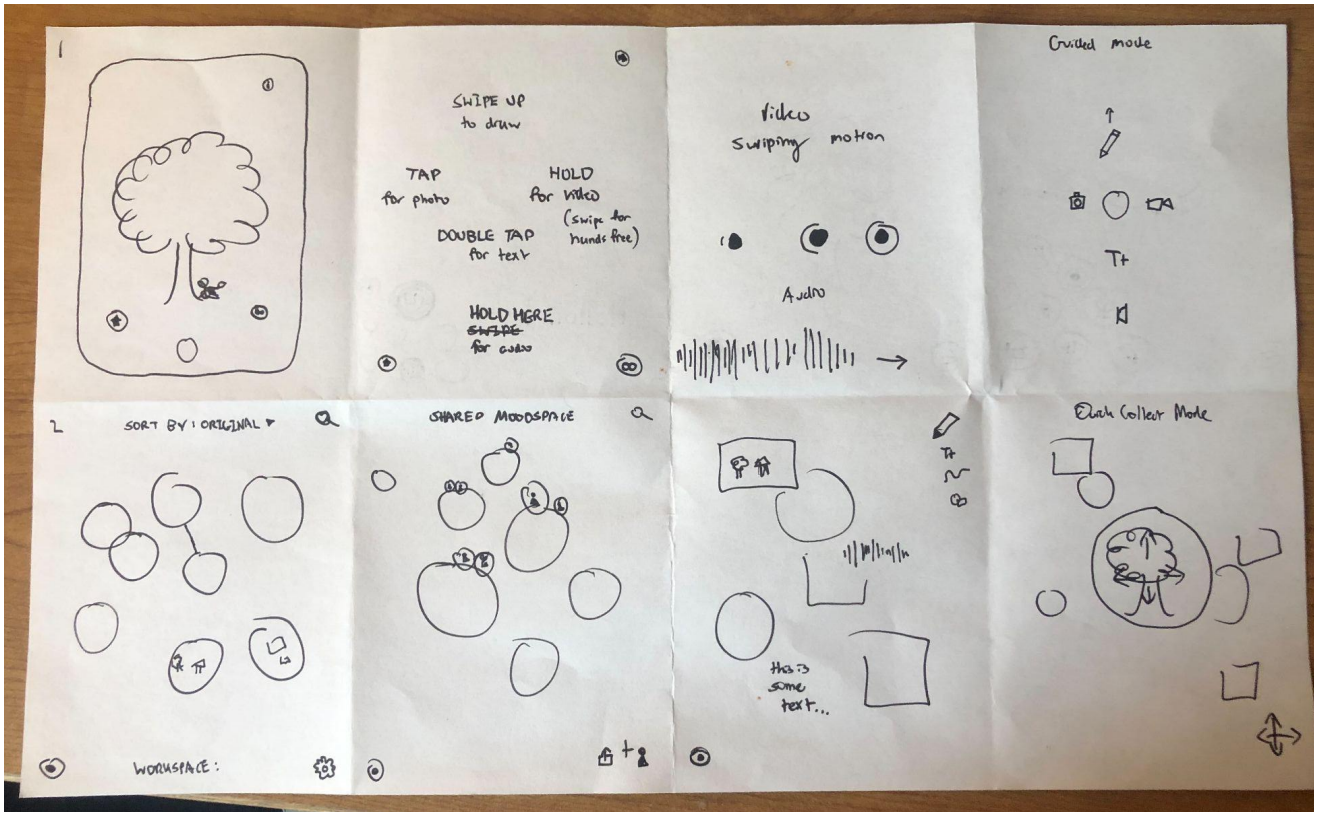


Figure 5: Sketches of an app idea that included sorting and viewing moodboards as a mood board itself, as well as a gesture-based UI for capturing.

UI Sketches

We converged to two form factors to draw out storyboards for: a wearable device and a mobile app.

1. Wearable

Figure 6: A more detailed version of the wearable concept, which captures and records multimedia automatically and syncs with a phone app to view moodboards.

Pros:

- A wearable is very fast and always available/ready
- Wearables are becoming more and more popular, especially for art
- Has the potential to be an artistic piece of fashion on its own

Cons:

- Jewelry can get in the way of certain art practices — for example, a cellist can't wear a bulky necklace
- Harder to observe and take photos and other visual media as the wearable itself has no screen
- Privacy issues — automatic capture could easily become intrusive
- A wearable gives you less control over capture

2. Phone App

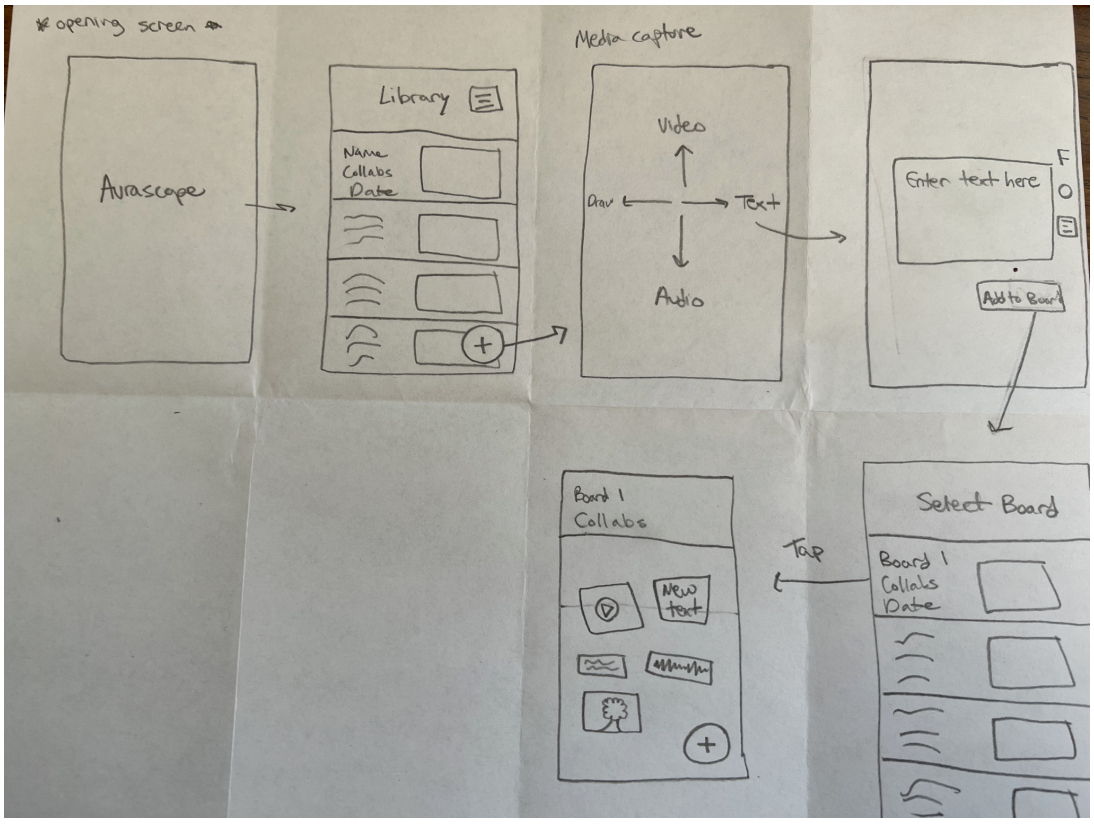


Figure 7: A more detailed app workflow sketch, including linked pages and task flows and a gestural-based capture screen.

Pros:

- Nearly everyone carries a phone with them
- The interface of media capture on a phone is familiar, and already used by many artists
- Wide variety of media to capture on modern phones.
- Phone media quality can be as good as professional equipment

Cons:

- Not as convenient to pull out to capture inspiration — requires more than one step to turn on and navigate to an app
- Not as discreet as a wearable
- Sharing and collaboration features are much easier to integrate — people already use phones to text and email

Final Selection: Phone App

In the end, we were excited to implement an app analogous to the combination of social media, notetaking and design mobile apps we were inspired by. We wanted something accessible; a wearable would have to be bought, whereas a phone is already carried by most artists — although we recognize even phones are still luxury items for many. Because of their universality, there is less of both a financial and learning barrier to using an app. Furthermore, we found the idea of a wearable to be too much of a potential invasion of privacy, and considered the ethical ramifications far too dangerous for us to be able to plan for and handle.

Selected Interface Design

Task flows

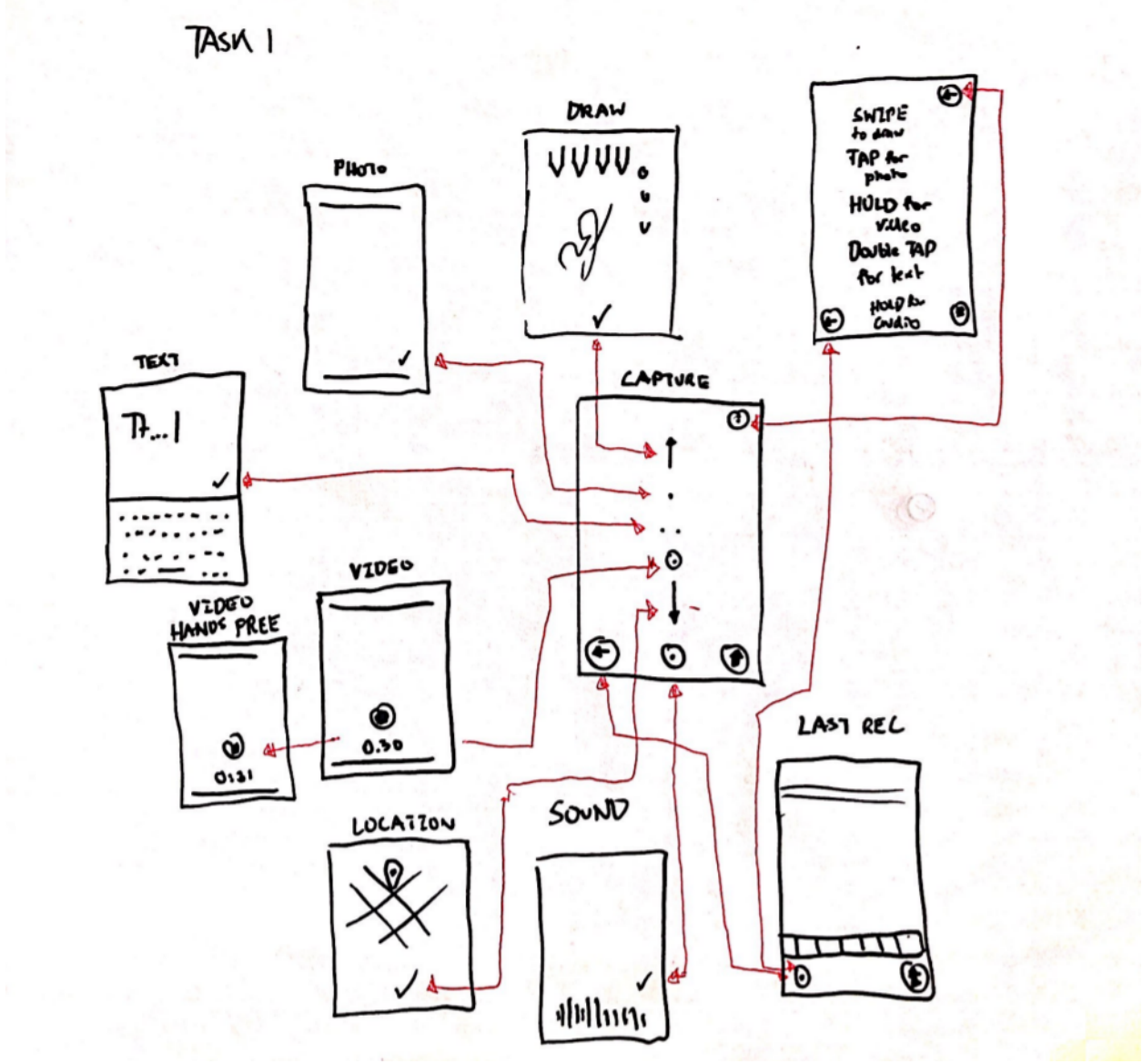


Figure 8: Simple task: Quickly capture and save audio, pictures, video, writing and drawings using the app

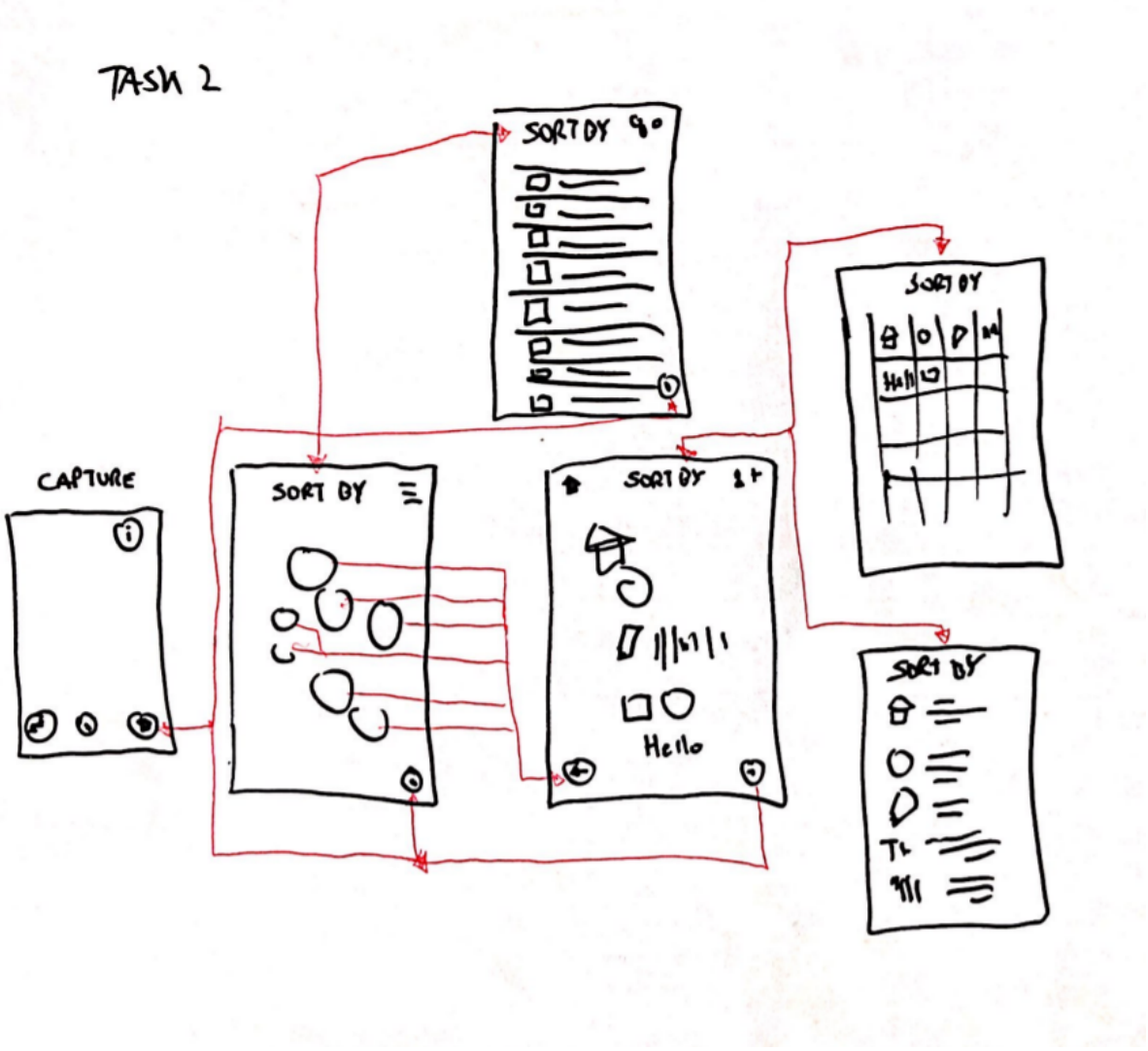


Figure 9: Moderate Task: View synthesized moodboard automatically generated from captured multimedia

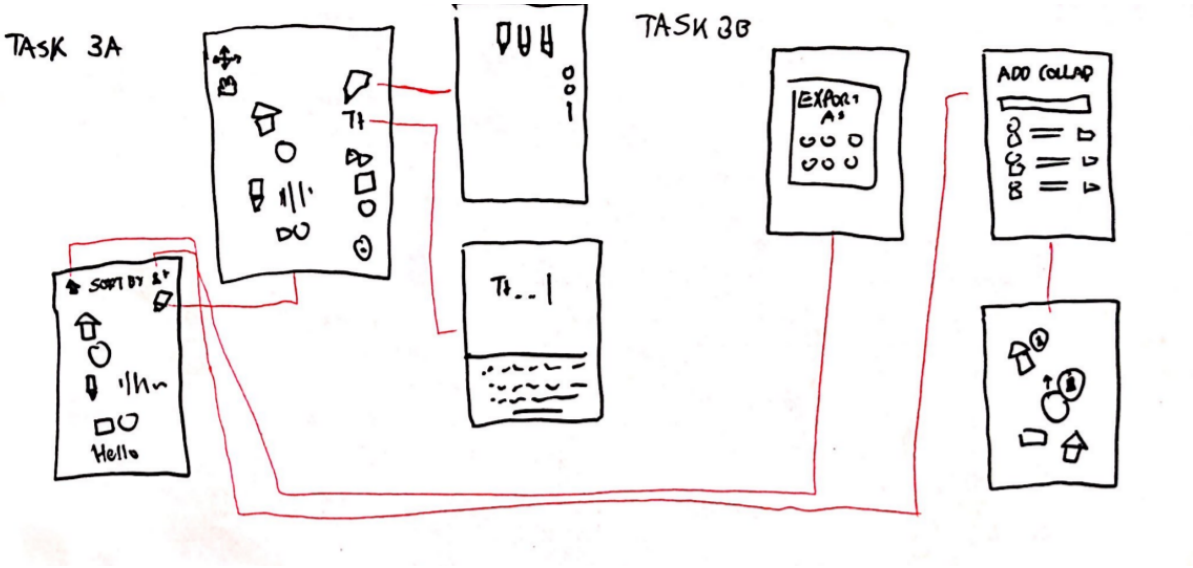


Figure 10: Complex Tasks: Customize moodboards using captured media and other creative tools, and share and collaborate on moodboards with other users

Lo-Fi Prototype

We opted to use Balsamiq for our prototype over refitting our existing sketches, mainly for the sake of simulating an image on our capture screen, but also for its cleanliness and interactability.

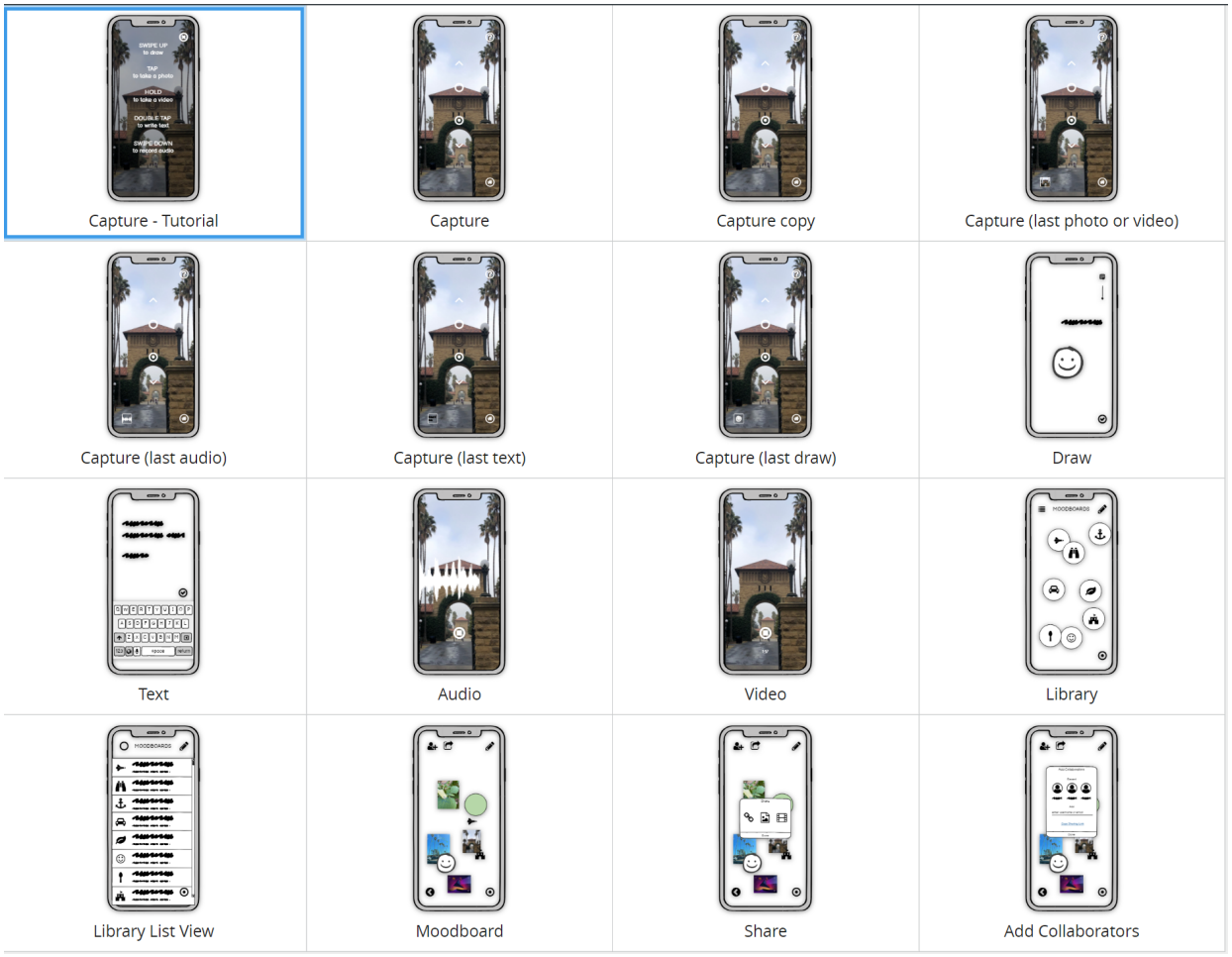


Figure 11: Overview of our Balsamiq mockups

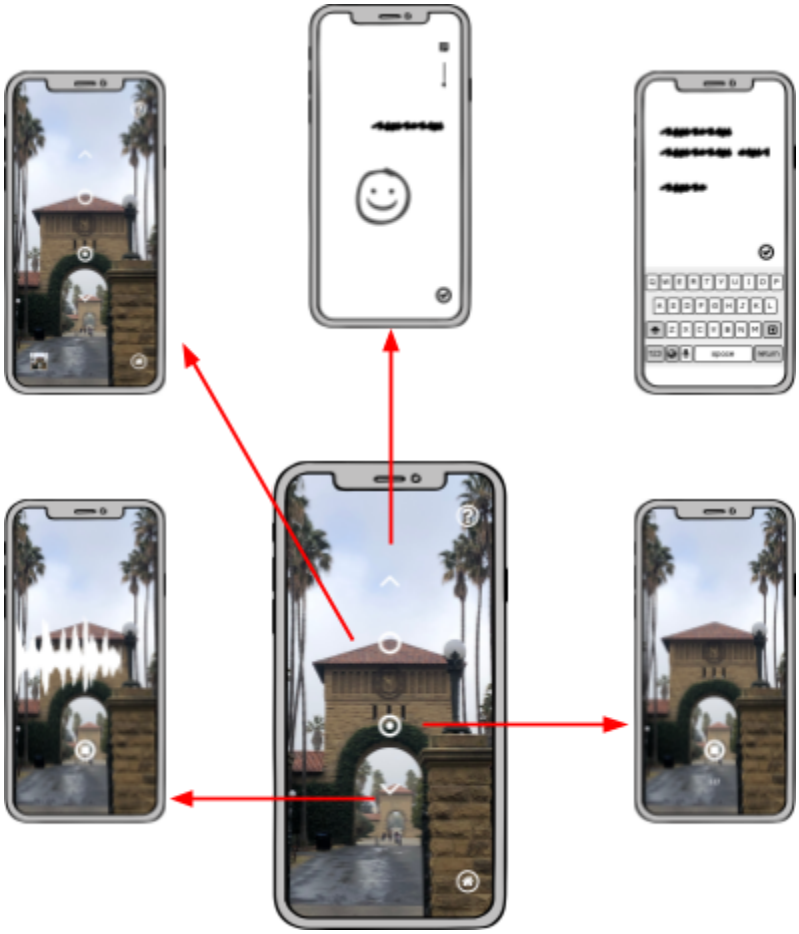


Figure 12: Simple task flow Balsamiq overview, simulating capture of media

Figure 13: Moderate task flow showing different ways of viewing moodboards, including a list of moodboards and a "moodboard" of moodboards, as well as an example of a moodboard itself.

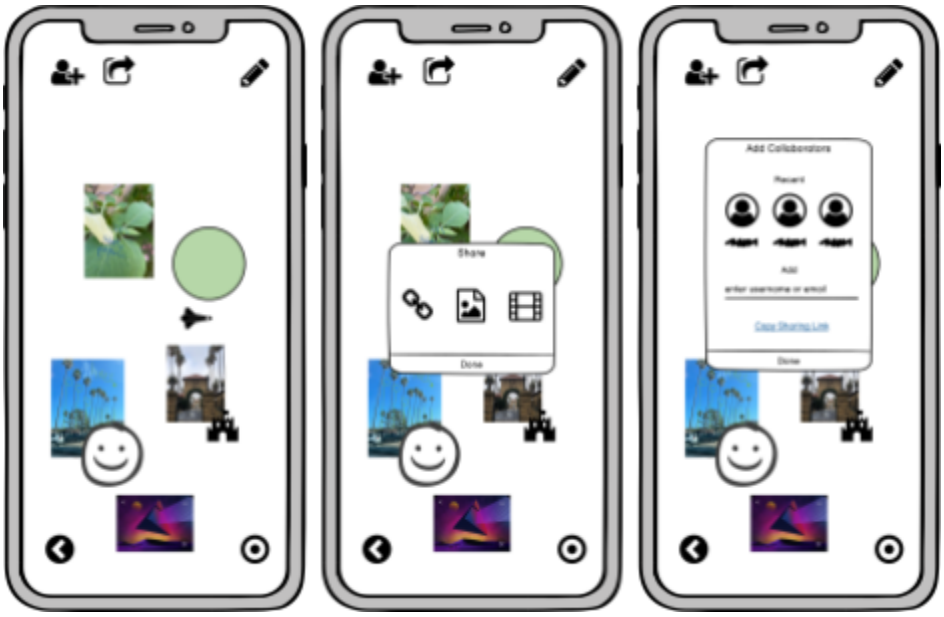


Figure 14: Complex task: sharing and collaborating a moodboard

Testing Methodology

Participants

We recruited three participants who were considered acquaintances by the researcher team to mitigate responder bias. Participants were not compensated for their time.

Our three participants ranged in age from 19 to 31 (median of 26). We interviewed two females and one man. Each individual had experience or interest in creative work. One is a Stanford student, one a BFA graduate in Industrial Design, and one a software engineer (with interest in photography and television). Two individuals identified as Asian and one as White.

Figure 15: Pixelated screenshots of our participants. We did not record one call due to technical issues with Zoom.

We identify our three participants [P16, P17, P18] by the nicknames "Dory," "Nil," and "May," respectively, based on their preference.

Environment

All participants were interviewed over a Zoom video call. Dory and Nil both shared their screens with us as they interacted with our low-fi prototype on Balsamiq. To better test the mobile phone characteristics of our prototype, we pivoted to using a Wizard-of-Oz mobile phone experience where May viewed a screen share of a researcher's screen and spoke out loud her thoughts and where she wanted to press a button on her phone. The researcher then pressed or held the buttons accordingly.

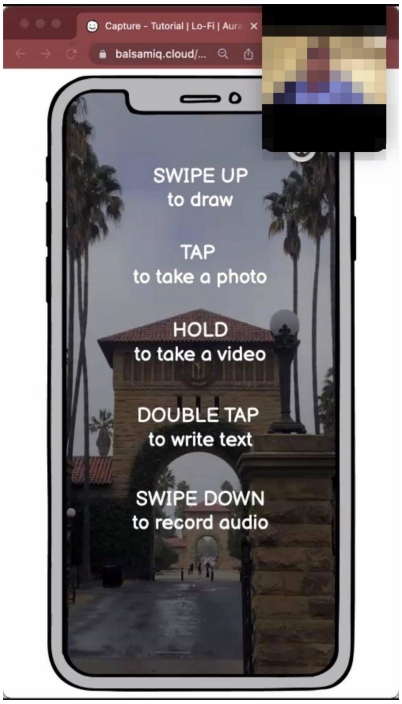


Figure 16: Screenshot from May's iPhone showing that a team member's Zoom video thumbnail is obscuring key elements of our user interface.

The image above illustrates a limitation of our testing strategy wherein key elements of our Balsamiq UI were obscured by our participant's iPhone's Zoom client.

Tasks

We asked our participants to walk through three tasks:

1. Capture and save audio from your environment using the app
2. View synthesized moodboards that were generated from your captured multimedia
3. Share and collaborate on your mood boards with other users

Usability Goals

We focused on three usability goals, each with their own metrics:

1. Efficient: perform tasks quickly

Our participants should be able to capture content quickly. To evaluate this goal, we counted the number of clicks or taps taken for each task. A lower count indicates more efficiency.

2. Pleasing: high user satisfaction

Our participants should find aesthetic and functional value in their app experience. Our first measurement tool was the Net Promoter Score from NICE SatMetrix which asks, "on a scale of 1-10 would you recommend this system to a friend?"

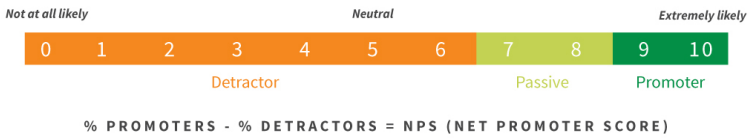


Figure 17: NICE SatMetix Net Promoter Score (NPS) scale

Only a score of 9-10 indicates that a person would actually recommend this system to a friend, a fair proxy for high user satisfaction.

Our second measure was the Task Load Index (TLX) survey developed by NASA. (See Appendix 2: NASA TLX Survey). This survey assesses the perceived demand of a task on a user along dimensions like mental demand, physical demand, and perceived performance. We adjusted the 21-point survey to a 10-point scale so we could administer the survey over Google Forms. (We administered the demographics and NASA TLX survey for P16 verbally but decided to pivot to a form because it could subject participants to expectancy bias and discomfort in answering personal questions.)

3. Learnable: faster the second time

Our users should be able to quickly master the basic tasks and interaction patterns of our app (such as getting to the share button within a few seconds). We measured the learnability of our app by measuring the start-to-finish time taken to complete each task. A lower time is better here.

Second, we were sure to write down every question that was asked of us during the interview sessions to understand gaps between the system's goals and the user's path to getting to that goal.

Procedure

We consolidated our procedures into one document with a clear script to ensure consistency amongst our participants. Briefly:

1. Facilitator gains consent to record the video call.
2. Facilitator asks the participant to fill out our standard demographics survey on Google Forms.
3. Facilitator walks through the basics of Balsamiq.
4. Facilitator asks the participant to screen share their Balsamiq webpage.
5. Facilitator asks the participant to show us how to "capture and save audio from your environment using the app" (Task 1) while thinking out loud.
6. Facilitator asks the participant to fill out the NASA TLX survey for Task 1
7. Facilitator repeats Steps 4-6 for Task 2 and Task 3.
8. Facilitator asks the NPS scale question, asks the participant for any remaining questions, and debriefs the participant.

For more details, please see Appendix 3: A5 Testing Procedures Script. The observer/note-taker team members logged critical incidents during the call.

Key & Other Test Measurements

Summarizing from the Usability Goals and Procedure sections above, our test measurements were:

1. Number of clicks or taps taken for each task
2. Start-to-finish time taken to complete each task
3. NASA TLX for each Task
4. Net Promoter Score
5. List of questions asked by participants of us

Team Member Roles

This A5 assignment description outlined four roles for an interview: facilitator, observer, note-taker, and computer. We assigned the role of Facilitator to a team member who did not recruit them. We consolidated the role of observer and note-taker, and used the computer role only for May.

Participant	Role			
	Facilitator	**Observer**	**Note-taker**	**Computer**
P16, "Dory"	Pramod	Katie	Katie	N/A
P17, "Nil"	Katie	Pramod	Pramod	N/A
P18, "May"	Katie	Pramod	Pramod	Pramod

Table 1: Division of roles amongst research team

Results

Time

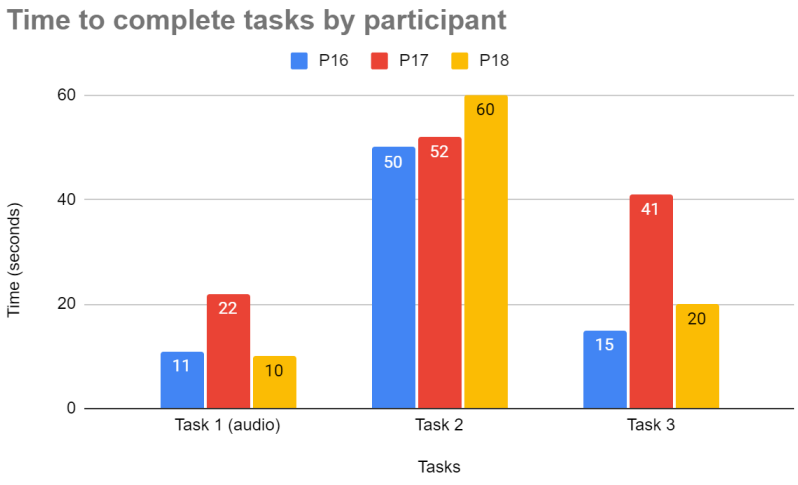


Figure 18: Approximate time it took for each participant to complete each task. Task 1 seemed to be the quickest to figure out, but task 2 and 3 were confusing for many.

Clicks

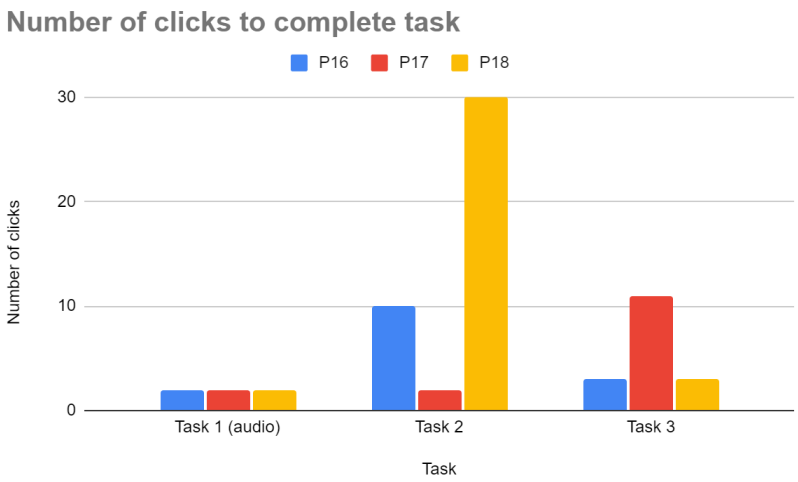


Figure 19: The counted number of clicks it took for each task. Task one only took 2 clicks for all participants, whereas tasks 2 and 3 took some more clicking to discover.

Task Workload

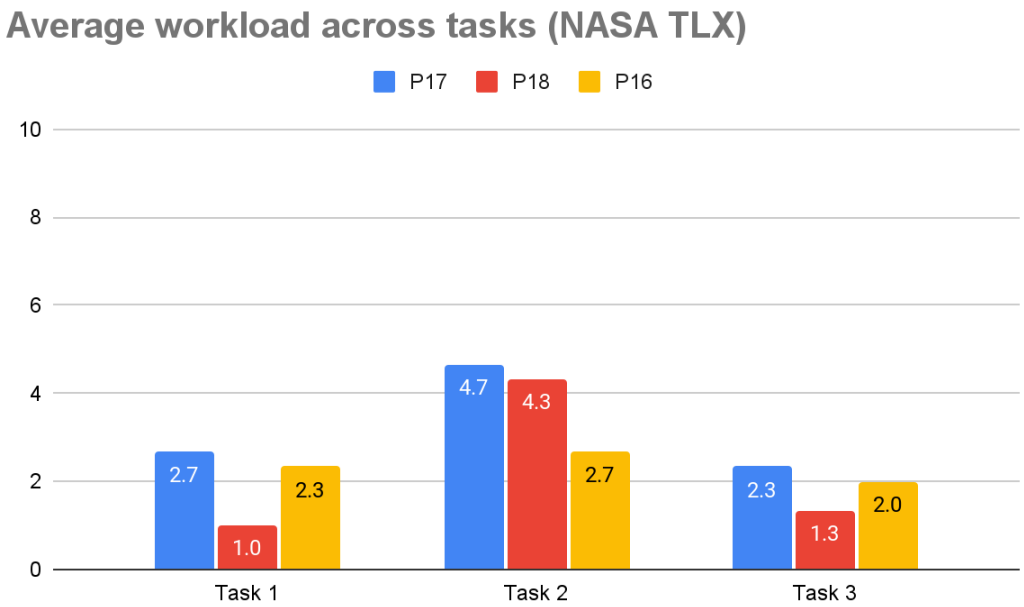


Figure 20: Taking the average of the six levels of demand in the NASA TLX survey gives us a general sense of workload across tasks. We observed low workload for all tasks, but task 2 might require some work.

Net Promoter Score (NPS)

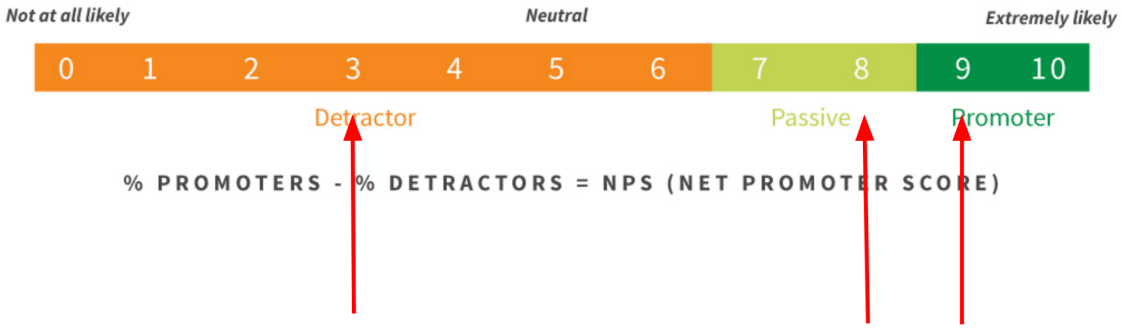


Figure 21: Only one of our participants rated our app highly.

Discussion

Quantitative Results

The number of clicks and time to complete each task was generally low for Task 1. Task 2 was more confusing in general, e.g. P18 couldn't find our home button because it was obscured by Zoom. P18's confusion revealed that our UI for Tasks 2 and 3 had the potential to be more confusing. Our workload measures indicate that workload was generally low amongst all tasks and that Task 2 required more demand. Our usability goals of being *pleasing* and *efficient* were only partially supported by our prototype tests.

We believe that Task 3 had lower (better) metrics because of a learning effect from completing Task 2 before Task 3. This difference supports our *learnable* usability goal.

Our NPS score suggests that our app does not have a "Promoter" characteristic indicating a neutral level of user satisfaction, i.e. a neutral *pleasing* goal.

Qualitative Results

P17, Nil, was most ambivalent towards our app experience and pointed out points of confusion in our app design. He commented that "I don't really know what I'm looking at." on the "bubbles screen" for Task 2 (viewing a synthesized mood board). Before embarking on Task 2, Nil commented "What am I doing again?" These quotes supported our conclusions from our quantitative results leading us to remove the 'bubbles' screen from our UI.

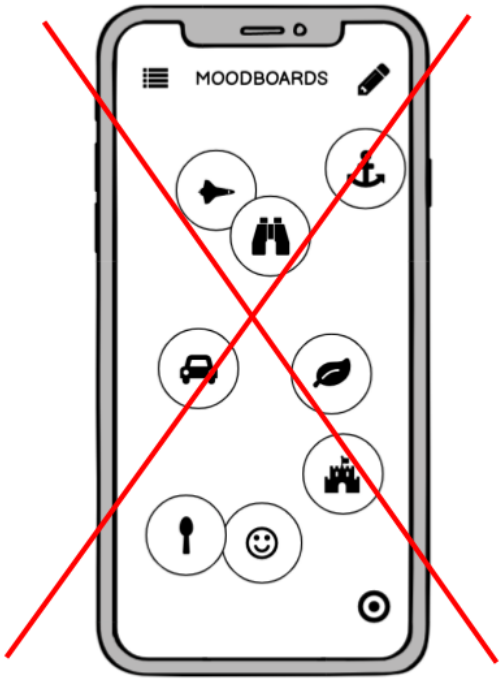


Figure 22: We removed the 'bubbles' screen following feedback from users.

P16 remarked that "even if [our app] wasn't intuitive it was easy to like figure [sic] out what was going on" supporting the neutral score she gave for our NPS survey. However, these issues could have been a result of our testing strategy. (P17 remarked "On a real phone would this be swiping?")

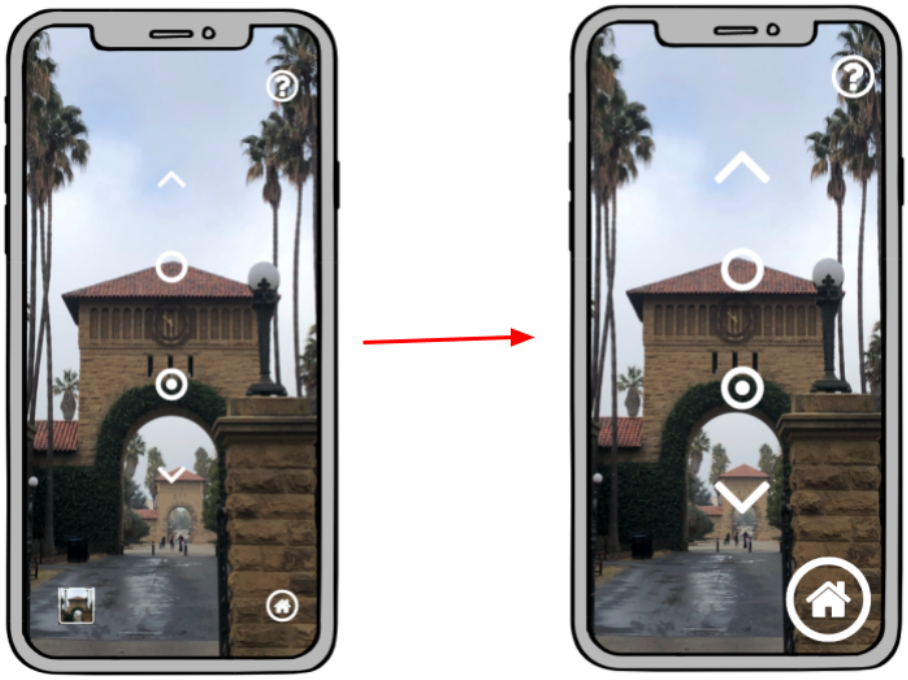


Figure 23: We resize the buttons for greater reachability and visibilabilty.

Appendix

Appendix 1: Consent Form

Consent Form

Aurascope's prototype is being produced as part of the coursework for Computer Science course CS 147 at Stanford University. Participants in the experimental evaluation of this prototype provide data that is used to evaluate and modify the interface of Aurascope. Data may be collected by interview, observation and questionnaire.

Participation in this experiment is voluntary. Participants may withdraw themselves and their data at any time without fear of consequences. Concerns about the experiment may be discussed with the researchers (Nathan, Katie, and Pramod) or with Professor James Landay, the instructor of CS 147:

James A. Landay
CS Department
Stanford University
650-498-8215
landay at stanford dot edu

Participant anonymity will be maintained by the separate storage of names from data. Data will only be identified by participant number. No identifying information about the participants will be available to anyone except the student researchers and their supervisors/teaching staff.

I hereby acknowledge that I have been given an opportunity to ask questions about the nature of the research and my participation in it. I give my consent to have data collected on my behavior and opinions in relation to the Aurascope's research. I understand that I may withdraw my permission at any time.

I give consent to be videotaped during this study:
___Yes ___No
I give consent to be audiotaped during this study:
___Yes ___No
I give consent for video or audio recordings from this study to be shown to people not directly involved with this research during/in class, seminars, reports, or scientific presentations.
___Yes ___No

Name [redacted]

Participant Number [redacted]

Date [redacted]

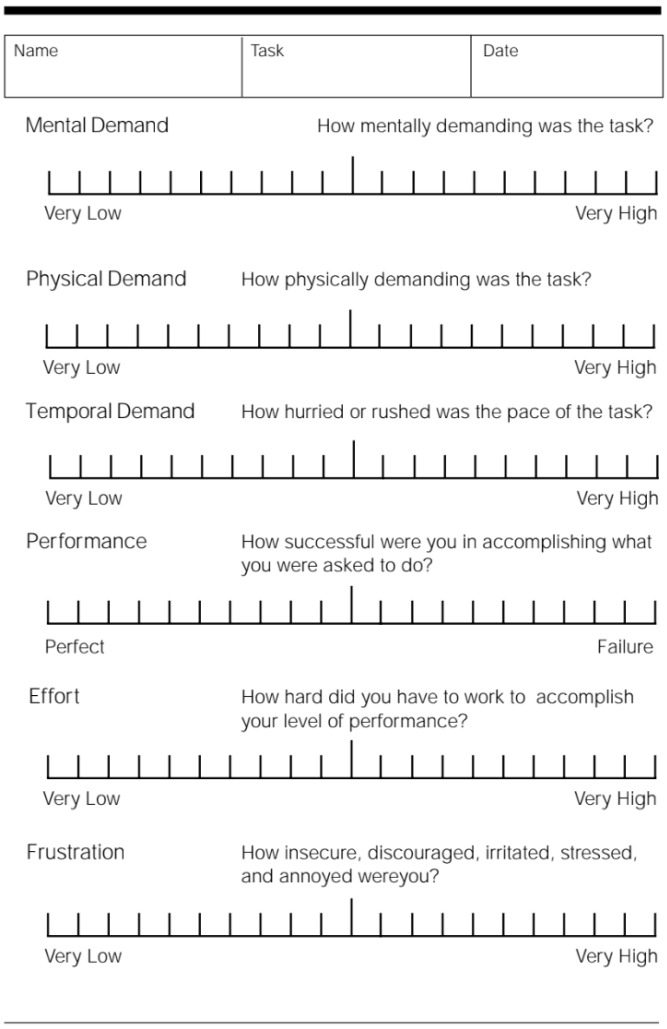
Signature _____

Appendix 2: NASA TLX Survey

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.



Appendix 3: A5 Testing Procedures Script

For remote testing:

- Participant opens Balsamiq link on their laptop: https://balsamiq.cloud/sl17z2a/p1bmk4q/r1F7E?f=N4IgUiBcAMA0IDkpxAYWfAMhkAhHAsjgFo4DSUA2gLoC%2BQA%3D
- Participant shares their screen with us
- We ask them to step through this Google Form as the study goes along: https://forms.gle/EP44s3gpNi3Mr5KNA

Researcher demos system: (see script)

Script:

1. Hello! Thank you for participating in our prototype evaluation session. We appreciate your time. First, we'd like to get a sense of who you are!
2. Do you mind if we record audio and video from this call for our future reference? Again, your identity or likeness will not be associated with you when we discuss our findings from this session.
3. Facilitator: Administer **not in-person** "Onboarding Survey" from 2022-03-12 - Final Report *[Ins2] Instrument 2: "Study Participant Onboarding Survey"*
4. We'd like you to interact with a prototype application we have created in Balsamiq, a low-fidelity prototype-creation tool that we can both access from the internet.
5. Facilitator: Share your screen and open the Balsamiq link.
6. Let's walk through an example of how you could interact with the system:
    a. On this screen [right], you can, for example, tap to record a video of what you are seeing as if you were holding up this phone on Stanford's campus.
    b. Here, we'll "X" out of the information page.
    c. Facilitator: Tap on "X" in the upper right corner of the screen.
    ~~d. For example, pressing the "target" icon [right] in the middle of the screen will start recording video.~~
    e. Facilitator: show interaction with the app.
    f. Any questions?
7. Great! Can you open up the Balsamiq link and share your screen?
8. Let's get you started with Task 1. We would like you to "capture and save audio from your environment using the app." Can you show us how you would do that? Please do remember to think out loud!
    a. Participant: opens and closes the audio screen from the app.
    b. Great, let's do that again, but now "capture pictures or video using the app."
    c. Participant: opens and closes the picture taking or video recording screen from the app.
    d. Great, once more, but now "capture drawings using the app."

    e. Participant: opens and closes the drawing screen from the app.
    f. Facilitator: Administer NASA TLX.
    g. Thank you! Any questions?
9. Let's move on to Task 2.
    a. We would like you to "view synthesized moodboards that were generated from your captured multimedia." Can you show us how you would do that? Please do remember to think out loud!
    b. Participant: eventually navigates to the moodboards screen.
    c. Facilitator: Administer NASA TLX.
    d. Thank you! Any questions?
10. Great, let's move on to Task 3.
    a. We would like you to "share and collaborate on your mood boards with other users." Can you show us how you would do that? Please do remember to think out loud!
    b. Facilitator: Administer NASA TLX.
    c. Thank you! Any questions?
11. Facilitator: Ask survey questions on NPS and friend recommendations (from below) & debrief them.

Appendix 4: Critical Incident Logs

Participant Number	Task Number	Severity Rating	Description
P16	Task 1	0	No problem. Body language and mouse movements suggest that this is an easy task.
P16	Task 2	1	Cosmetic problem. "Oh wow" comment suggests being overwhelmed when clicking on the 'menu hamburger.'
P16	Task 3	0	No problem. "I think this is it?" comment when reaching the share buttons in the prototype.
P17	Task 1	3	Major usability problem. "On a real phone would this be swiping?" comment suggests that the gesture UI is not clear and/or our testing method is flawed.
P17	Task 2	3	Major usability problem. "I don't really know what I'm looking at" comment on the bubbles screen and a look of confusion or overwhelmedness.
P17	Tasks 2-3	3	Majority usability problem. Appears confused for much of the time taking the task.
P18	Task 2	4	Usability catastrophe. P18 spent over one minute and dozens of clicks to try to complete Task 2. She did so because the Home Button was obscured by the Zoom client on her phone.
P18	Task 2	2	Minor usability problem. Visible confusion between 'tapping' and 'holding' actions leading to much more time spent on Task 2.
P18	Task 3	1	Cosmetic problem. Seems overwhelmed when seeing the menu from the 'hamburger' button.