

ASSIGNMENT M5

Kalpana Baheti

kbaheti3@gatech.edu

Abstract—This series of qualitative and quantitative studies will cover different stages of interface understanding, design, and evaluation for a mood-reading music player which has its controls governed by an app on a phone. The music player is meant for a playroom at home or at a school where children spend significant amount of time. Music has a long-standing effect on children's cognitive and emotional capacities; this device will be responsible for deciding what to play and when with a desired ambience/accomplishment as a goal. The music auto-playing system has two components – the environment-reading music device and the phone app which controls this device. We will be studying the interaction of the user (the educator/parent/guardian) with the app's control panel given fixed assumptions on the manner in which the settings will regulate the environment-reading device's input and output.



Figure 0 - My Original Sketch

Catching up on previous information...

The key demographic for the usage of the app control panel are parents, teachers, and guardians of children aged 5 to 13, who have average to better than average cognitive ability and belief in the effect of music on child development, who belong to well-versed English tongued population, who have the monetary capacity to purchase such a product set, and who have sound haptic, auditory, and visual abilities. **We will performing empirical and predictive evaluation.**

Note on Prototypes Used –

Version 1 Prototype –



Figure 1 – Basic Control Panel

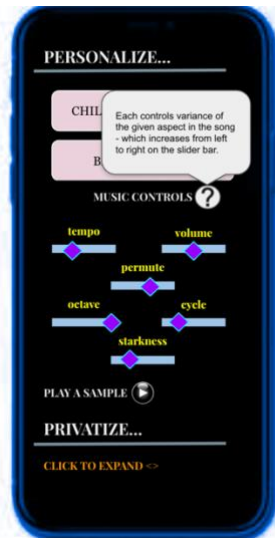


Figure 2 – Click Question Mark

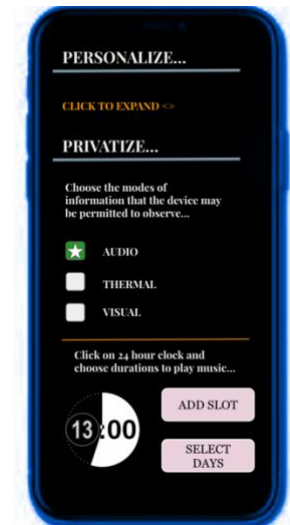


Figure 3 – Expand Prioritize

Version 2 Prototype –

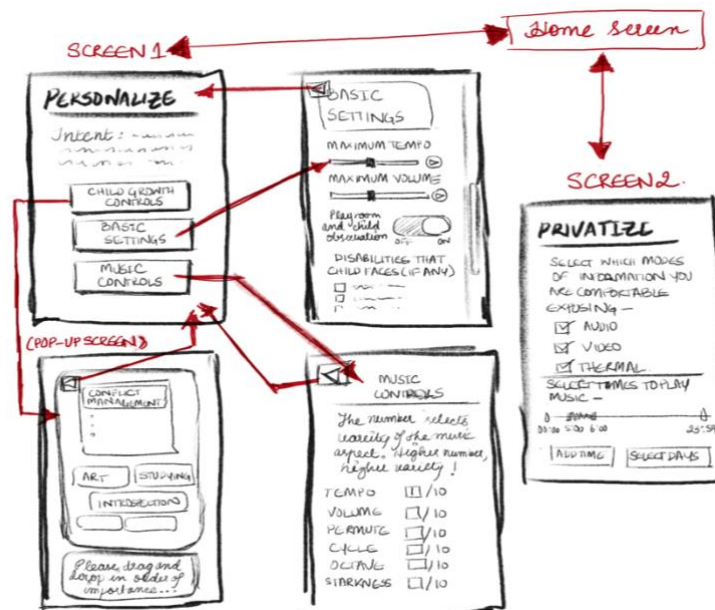


Figure 4 – Card Prototype

EMPIRICAL EVALUATION

Prototype Version Selection and Definition –

Version 1 Controls:

Personalization: Unordered options for child growth goals (equal weightage to each), adjusted (as per description above) selection panel for base constraints, *there will be no selection for advanced musical aspects (just tempo and volume included in base).*

Privatization: Default set to audio information intake and afternoon time, option lists and time interval bars used for this respectively.

Version 2 Controls:

Personalization: Ordered options for child growth goals, adjusted (as per description above) selection panel for base constraints, continuous horizontal sliders – provision of automatic examples upon setting music aspect.

Privatization: Default set to audio information intake and all-day observation, option lists and time interval bars used for this respectively.

Data Recorded – Post-Event Close-Ended Questions

- a. Metric 1 Scoring – Fulfilment of Goal
- b. Metric 2 Scoring – Ease of Understanding
- c. Metric 3 Scoring – Ease of Navigation
- d. Metric 4 Scoring – Comfort with Information Asked
- e. Metric 5 Scoring – Aesthetic Appeal

Testing Process – Post-Event | Synchronous | Live Demonstration | Individual Feedback – no deviations, and everything went smoothly.

Experiment Design and Definition of Null/Alternate Hypothesis –

A **randomly selected half within-group** manner (collected from halves of each sub-group of our key demographic) will be assigned to version 1 and the other half to version 2. I'm keeping it simple, **no ordered selection** here.

Regarding method of usage and data collected; since this is being clubbed with qualitative evaluation observation stage, all this information is mentioned there.

We have 5 metrics listed. **For each of metric, we receive a score – ordinal data.**

Hence, we will use the **Chi-Squared Test** for each of our metrics.

H_{NULL} for Metric *i* → If both distributions for Metric *i* are **equal**

H_{ALTERNATIVE} for Metric *i* → If Distributions for Metric *i* are **unequal**

Raw Scores out of 5 – Found in the Appendix: Conducted for...

30 People, 2 Groups (of 15), 5 Metrics, Row Order is Version 1 then Version 2

Results for Chi-Square Test –

Metric 1 (Goal Fulfillment) – Both Versions Cleared!

Chi-square statistic: 5.173421779988943

P-value: 0.983249719047807

Fail to reject the null hypothesis. Equal Distributions!

Metric 2 (Ease of Understanding) – Version 2 Better!

Chi-square statistic: 5.235687633262262

P-value: 0.9822546147307654

Reject the null hypothesis. Unequal Distributions!

Metric 3 (Ease of Navigation) – Version 2 Better!

Chi-square statistic: 1.740215937200262

P-value: 0.9999648421997571

Reject the null hypothesis. Unequal Distributions!

Metric 4 (Comfort with Information Asked) – Both Versions Cleared!

Chi-square statistic: 6.570158730158729

P-value: 0.9500

Fail to reject the null hypothesis. Equal Distributions!

Metric 5 (Aesthetic Appeal) – Version 1 Better!

Chi-square statistic: 1.6875000000000002

P-value: 0.999971000188696

Reject the null hypothesis. Unequal Distributions!

Did they match what I expected...

Well, yes. Version 2 had far fewer controls because it did not contain the fine-grain music controls to trade-off bias and variance of tempo, volume, starkness,

cycle, and so on, in the music. That was a part only a core segment of music and education aficionados really took to.

Version 1 on the other hand, really was designed prettier and more compact. The pastel board colors made it pleasurable to look at and the compact nature of the app controls (no new pages, like the card prototype, all in one page) created a feeling of getting done with the task quickly.

If not, why do I think the results were not as expected...

Nothing went as I didn't expect, per se. But I was disappointed that the overall score of the Version 1 is lower due to the extra controls (Version 2 had 7, Version 1 had 12). The reason is that version 1 is truly capable of giving a far more memorable experience to children. I think this experiment reveals a lot regarding how much awareness parents have on this section of science, and if there is awareness, how much energy they have to build this experience.

If so, do I believe the results are because of real differences or because of lurking variables or experimental errors...

There might've been randomly occurring bias in the two groups, and there was a shortage of enthusiasts (which are naturally in shortage across the whole population out there, and not just in my sample). I would also increase the total count of voters to 100 as I knew that many willing people who fit the target population.

What changes would I make...

I would take the best points of the first version and incorporate it into the second – keep the color board the same, remove the page jumping and keep expandable tabs, and remove the six advanced controls while still keeping the 'Play a Sample' option from Version 1.

PREDICTIVE EVALUATION

I will be describing a **GOMS model** which covers –

1. Initial State
2. Selection Rules

3. Version 1 and Version 2 (Different Methods) Operator Pathways
4. Time Estimated per Operator
5. Goal State

Note 1 – The (s) stands for seconds.

Note 2 – The operators available are listed in the GOMS model below, it is the EXACT set – as in, no operator got left out, and no excess operators were needed and were deficit in the model.

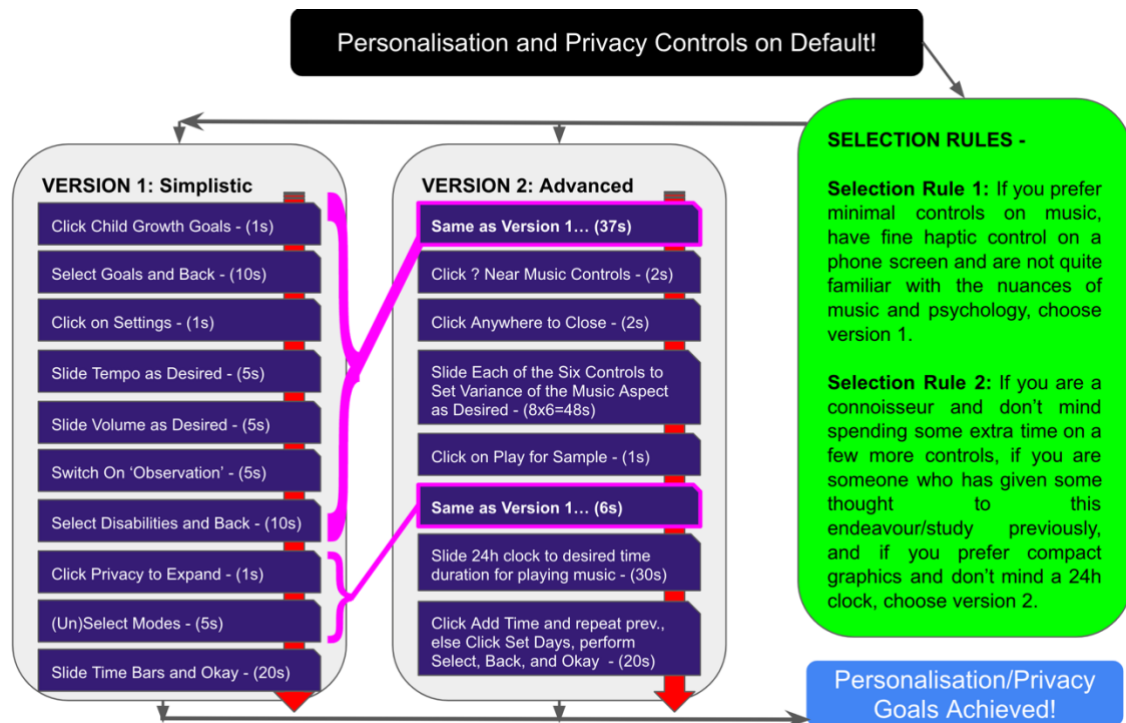


Figure 2 - GOMS Model for Version 1 and 2

Could it be made more efficient?

The time bar would be a better alternative to the 24-hour clock since that may be tedious to move with a finger, even if it's compact. Also, the default time duration may be set to afternoon times.

I think another easy way to get this whole sequence of operations performed easier is to connect the commands with Siri or whatever chat bot android OS

uses. This would especially help with the time duration setting. That seems to create the most inertia of action.

What would the trade-offs be?

The trade-off with using a chat bot comes in two parts –

- 1) There is significant engineering required to connect an app with Siri.
- 2) The commands become probabilistic (instead of being deterministic as it is in this case) and may enter wrong input or no input at all for choice.

What could be simplified, and what other parts might become more complex as a result?

One of the parts that could be simplified is converting the slider bar inputs to numerical scores out of 10. Problem there is that it's not natural to think of volume and tempo in terms of a score out of a total number, we just increase/decrease and check.

Of all the operators, the slider bar presents the most energy-consuming haptic experience (place exactly on puck, press, move to exact location, let go, sigh!). It is possible to make this slightly easier by drawing the sliding puck bigger so placing finger on it becomes easier – however in that case, the space for rest of the surrounding items reduces in turn.

EVALUATION SUMMARY

Describing Additional Need-Finding...

The interface, on both the personalization and the privacy front, I believe have been assessed well. The reason I'm convinced of this is because for some of the most important tellers of interaction experience quality, I have started at minimum and maximum options in these aspects and objectively (through user-in-loop) worked my way to the optimum configuration.

However, there is one aspect that I feel warrants greater attention. That's the interest and belief that parents and teachers have in this kind of luxury to begin with. I specified a set of criteria in M1 for the target population which covered income and literacy as well. But I don't believe that's enough to find out who

truly appreciates such a product. Literacy isn't enough, introspective individuals well-read in different fields value such endeavors far more. Income levels aren't enough if the spending population does not have the education mentioned. I would spend resources quantifying the connection between these factors and the trust in the product, and then identify what the true market for this product is.

Evaluation Regarding Additional Design Alternatives...

A whole different design alternative does not seem needed. The changes would be to keep the color board the same from Version 1, remove the page jumping and keep expandable tabs, and remove the six advanced controls while still keeping the 'Play a Sample' option from Version 1. The rest follows Version 2.

There also was a suggestion in the participant need finding conducted in M2, where there was promise of adding a feature where a sample song is uploaded, and the new music is inspired by that. That may be worth exploring.

Brainstorming Regarding Next Steps...

We used a wireframe and card-based prototype at this stage. For the next level of fidelity, constructing a simple backend for these front-end prototypes would be quite telling. There might have been some disruption of flow while I handed new cards or changed wireframe slides for the users. The backend automation will fix that.

Next Evaluation Steps...

Empirical evaluation was used in this evaluation stage. For the next iteration, I would use first use qualitative evaluation to narrow in on the differences in the empirical evaluations gained in this iteration – particularly, metric, 2, 3 and 5. And then follow it up by one last round of empirical evaluation on the specific points – where there's only one difference between two versions. That should be very directly informative.

REFERENCES –

1. The Ed Discussion lectures and extra readings.
2. My friends, acquaintances and teachers from my old school who contributed to the empirical evaluation.

APPENDIX – RAW EMPRICAL SCORES OUT OF 5

30 People, 2 Groups (of 15), 5 Metrics, Row Order is Version 1 then Version 2

Metric 1 (Goal Fulfillment) –

3	5	5	5	5	5	2	3	5	5	5	5	4	5	5
3	3	3	4	4	5	3	3	3	2	3	1	3	1	4

Metric 2 (Ease of Understanding) –

3	3	3	3	3	3	3	1	5	5	5	3	3	3	4
4	4	4	5	5	5	5	5	3	3	5	5	5	5	4

Metric 3 (Ease of Navigation) –

4	4	5	4	5	4	3	5	5	5	5	3	3	3	4
5	5	5	5	5	5	5	5	5	3	5	5	4	4	5

Metric 4 (Comfort with Information Asked on App) –

3	3	4	5	4	5	4	3	1	2	1	2	5	4	4
2	1	2	5	5	5	5	4	4	4	2	5	4	3	3

Metric 5 (Aesthetic Appeal) –

3	3	4	4	4	4	3	4	5	5	5	3	3	5	5
3	3	3	3	3	3	3	3	5	3	3	3	3	2	5

