ASSIGNMENT M4

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.**

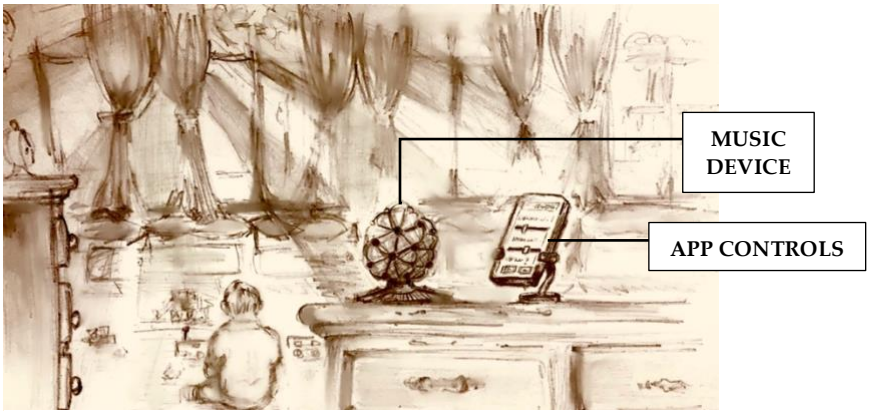

Figure 0 - My Original Sketch

Note: I will be using either the card-based prototype or the wireframe prototype as a presented item to the user. The wireframe is essentially the card protype completed in its aesthetic form and ready with buttons and different user pathways to take.

QUALITATIVE EVALUATION

Selection of Qualitative Method –

1. **Live Demonstration** – My app control panel, and the app itself (in terms of its purpose) has not been come across before, hence prior experience is ruled out here.

2. **Synchronous** – I will be sitting with individuals separately and observing (*notetaking*) how they use the app. I will not be sitting in front of them and breathing down their neck, but rather somewhere to the side or angled behind them with quite some distance, so they get a feeling of space and freedom.

3. **Two Versions** – While the previous paper covered three versions described, building wireframes takes time, and hence I will be sticking with two alternatives. I would like to keep this simple, so I will interview a greater number of candidates from the key demographic, but each half will be shown only one of the two prototypes.

4. **Post-Event Protocol** – I will be using a post-event protocol where after the user finishes a pathway from start state to goal state using the app's control panel, I shall receive summative feedback on it.

5. **Individual Feedback** – The trial and feedback sessions will be conducted individually and not in groups.

Evaluation Plan –

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.

1. **The Participants** – 10 parents of the key demographic, 5 elder siblings of the key demographic, and 5 teachers of the key demographic.

2. **Recruitment** – The participants will be gathered through my parents' acquaintances who fit the key demographic, my friends who are elder siblings who fit the key demographic, and teachers from my school from the International Baccalaureate and Cambridge sub-schools (where such aspects are given importance) – again part of the key demographic.
3. **Location of Evaluation** – Any room in the house/school where the participants may be comfortably seated and able to hold a phone and have enough time to consider what they'd like to choose – this is not a time-bound task.
4. **Recording** – There will be data recorded by a subtle observer (me), which I will cover in the following section. The recording will happen while the user is operating on the prototype.
5. **Post-Event Protocol Prompting** – At the end of the session, there will be a few open-ended questions, and a few close-ended questions.

Very Important Notes –

The data collected in this process is also going to be used for the empirical study. The candidate population accessible in the given duration is limited and I would need first-experience data. Hence, this exposure of the prototype will not be repeated.

Another point is that we are exposing two versions of the prototype – but they have (not many differences but) more than one difference. Hence, I am going to use the empirical study to determine which one is better on the whole and use the qualitative evaluation to understand '*why*' one prototype is better – in terms of parts of music personalization and child privacy.

Evaluation Content –

1. **Data Recorded 1 –** Notetaking During Observation
    a. Click path
    b. Time for each click
    c. Long pauses (also whether looking at screen or not)
2. **Data Recorded 2 –** 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

3. **Post-Event Prompts and Questions**
    a. What are your thoughts on the efficacy of these controls?
    b. What were the difficult parts to comprehend (if any)?
    c. What were the tedious parts to navigate across (if any)?
    d. What feedback would you like to share regarding privacy?
    e. Kindly share your thoughts on the looks and feel of the app.
    f. Anything else you'd like to share?

EMPIRICAL EVALUATION

Prototype Version Selection and Definition –

I will be presenting wireframes for both these versions, and as I mentioned earlier, this has more than one significant difference, hence we will only use this to find absolute scores – the reasons behind it will be deduced from the notetaking, and post-event qualitative protocol.

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.

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**

If distributions are unequal for a certain metric, we shall move to the qualitative and predictive analysis to see what causes the difference and reiterate.

Possible Lurking Variables –

1. **Randomly occurring bias in halved groups -** The group assigned to a more detailed version of the prototype (a connoisseur model of the app) may have participants who are very thoughtful and passionate about this or are on the other end where they see the value in this as a gift to their child but don't wish to experiment. Both these ends will bias results.

2. **Missing 'Feel' element –** This is not exactly how a prototype should be since while they will have visual stimulus, and even perhaps an audio enhancement, pressure-based haptics is missing – which will create some difference from the real feel for the real-world app control panel.

3. **Lack of segregation of impact –** While we're calculating based on an overall score for each of the five metrics, we're not able to assess which exact point led to the score in what way – however, this we're solving using other evaluation methods – though *not* quantitatively. This means

that if a metric showed similar distributions, they might still be great/bad at different junctures of the usage pathway.

PREDICTIVE EVALUATION

While this is a new interface, it is akin to interfaces used previously and commonly by most people (such as the Control Settings on Apple and Android phones), hence I will be describing a **GOMS model** which covers –

1. Initial State
2. Selection Rules
3. Version 1 and Version 2 Operator Pathways and Timings
4. 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. This covers the M5 part as well!

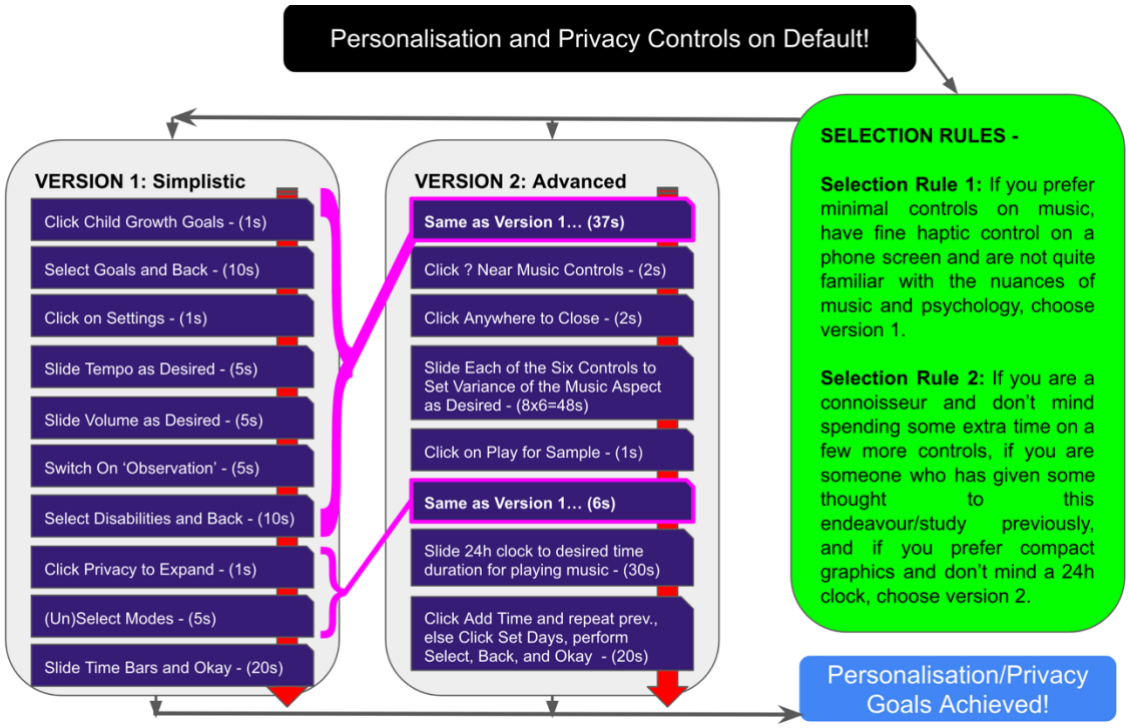

Figure 1 - GOMS Model for Version 1 and 2

PREPARING TO EXECUTE

Selection of Evaluation Methods –

I will be using qualitative and empirical evaluation methods.

Explanation –

1. The qualitative method covers a lot of relevant knowledge required in one sitting with each participant, that the quantitative methods would take a long time to cover piece by piece. It also gives a chance to the user to express diverse open-ended thoughts.
2. The empirical method gives a more concrete result that is scientifically verified, and it is possible to perform this in reasonable time.

I won't be pursuing predictive evaluation because I feel it is further away from reality than the other two and I would not be able to trust the results.

REFERENCES –

1. The Ed Discussion lectures and extra readings.
2. Assignment M1, M2, M3 intermediate results.