This document outlines potential experiment designs for the next spherical study.

Hypotheses:
[H1] A rested approach results in better interaction than unrested. *not trying to answer this one
[H2] Modeling the user's input space while rested results in a better interaction than a rested approach without modeling the input space.
[H3] A better model of the input space results in a better interaction
[H4] Gestural interaction will exhibit performance improvements over time. *identified during pilot
*we are really only attempting to answer H2, and H3 during this aspect of the study

1. AAA, BBB, CCC
    a. Participant calibrates once, and uses each model for 3 tasks. Once the participant is finished with all 3 tasks, the model is toggle-switched to the next one until all models have been evaluated 3 times.
    b. Advantages
        i. no recalibration after 1st
        ii. can identify learning effects in post-hoc of each model
        iii.
    c. Disadvantages
        i. Order seems to highly influence results.
2. ABC, BCA, CAB
    a. Participant calibrates at the beginning of each of the 3 rounds. Once finished the participant uses the calibration in 3 tasks where in between each task the model is toggle-switched.
    b. Advantages
        i. Order matters very little
        ii. if one calibration is bad there are others to fall back on.
        iii. literature backing from previous papers
    c. Disadvantages
        i. participants may learn the generic calibration rather than the differences between each model

Additional items:
consider incorporating ergonomic elbow rest
consider evaluating whether tilting the leap to enlarge the gestural space results in better performance, could be a small contribution. *not really worth focusing on
Spherical model is performing better but still slightly under planar from initial findings.

Alvin Comments:

How many devices you want to test, do you want to include the mouse? if yes: why, if no: why

how many participants do you need?

and correspondingly, how much would it cost you in terms of time, money, and statistical power

TODO
debug spherical model
balance target widths
use circular targets [now a setting]
construct new experimental orders that remove visual search time
Pilots
Experiment Script
Extra-Credit solicitation
Rewrite related works section
read UIST'14 papers

Paper Redesign:
We will be using the submission to CHI'15 as user study 1 for this paper. The key aspect of this study is that it ran 5 days using 5 participants to investigate learning effects. We then wanted to answer the modelling questions H2, and H3. We will be conducting user study 2 to investigate these questions. We will be addressing many concerns from the last paper including the sample size, visual search time in task, unbalanced target widths, related works criticism, excessive hover time, and possibly upping the index of difficulty of the task itself.

Summarised issues of last version:

CHI'15 Response

META
N too small (addressed)
Power reported incorrectly (conceded)
ISO Differences
- Randomness introduces visual search time (addressed?)
- Circular targets should have been used for consistent regardless of angle (addressed)
- Unbalanced target widths 15 20 20 20 15

Hover/dwell not ideal choice could increase fatigue (addressed)
NASA TLX for qualitative data should have been used (conceded)
Not reporting normality and variance assumptions of the ANOVA (conceded)
Hypothesis interpretations H2, H3

H4 not supported with statistical differences (addressed)
Not including reference (addressed)

> Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. CHI '14.

Benchmarks shouldn't be concerned with touchpad, unrested condition unnecessary as it was already shown in the WIP (not addressed or conceded)

Rev 1
Cite the above reference
Too much focus on fits and not enough on current state of gestures
No need for touchpad
Corner selection is not an appropriate reason to abandon the ISO standard
N too small
Power incorrectly calculated
Fatigue should be in NASA TLX
No normality and variance assumptions
H2 cannot be accepted
H3 should be rejected
H4 should have an associated statistical test

Rev 2
Reconfirming h1 is not enough
H2 is based on TRE and not missed targets
Failure to confirm H3 takes a large contribution away from the paper
H4 not confirmed with statistical tests and is not very surprising given well known benefits of practice for new techniques
The above gestural reference should be incorporated and the fatigue model should be used for analysis
Related works should again focus on recent gesture work not foundations
N too small
Hover is not ideal, slow, increases fatigue
Randomness of the target placement incorporates visual search time
Targets should be circular to control for direction of cursor entrance
" Effective ID could not have been between 200 to 216 bits. The range of ID in the experiment was 1 to 4.5 bits. Was effective ID only 2.0 to 2.16 bits? This is a very small range and any regression model to calculate throughput would not have much support."
No significance test of performance over 5 days
Should not report descriptive statistics when no statistical significance
"The touch pad can't possibly be 131 x 12 x 16 mm." should actually be 131 x 129 x 16 mm.

Rev 3
Confusing writing

N is too small, cant have such a high power
N should be larger for fatigue studies, statistics like throughput aren't as relevant here rather should focus on fatigue
Hover is notoriously known for being tiresome to hold, and no discussion of this
Table 1 is not clear
Table 3 comes before table 2
Describe raw data further
Calibration boundaries need further description