Your manuscript entitled "Virtual Reality and Haptic Control Interface supports Motor Skill Training with focus on Traumatic Brain Injury Recovery" which you submitted to Assistive Technology, has been reviewed. The reviewer comments are included at the bottom of this letter, along with those of the editor who coordinated the review of your paper.

I regret to inform you that the reviewers have raised serious concerns, and therefore your paper cannot be accepted for publication in Assistive Technology. However since the reviewers do find some merit in the paper, I would be willing to reconsider if you wish to undertake major revisions and re-submit, addressing the referees’ concerns.

Please note that resubmitting your manuscript does not guarantee eventual acceptance, and that your resubmission will be subject to re-review before a decision is rendered.

You will be unable to make your revisions on the originally submitted version of your manuscript. Instead, revise your manuscript using a word processing program and save it on your computer.

Once you have revised your manuscript, go to http://mc.manuscriptcentral.com/uaty and login to your Author Center. Click on "Manuscripts with Decisions," and then click on "Create a Resubmission" located next to the manuscript number. Then, follow the steps for resubmitting your manuscript.

Because we are trying to facilitate timely publication of manuscripts submitted to Assistive Technology, your revised manuscript should be uploaded as soon as possible. If it is not possible for you to submit your revision within a reasonable amount of time, we will consider your paper as a new submission.

I look forward to a resubmission.

Sincerely,
Dr Cooper
Editor in Chief, Assistive Technology
rcopper@pitt.edu

Reviewer(s)' Comments to Author:

Reviewer: 1

Recommendation: Major Revision

Comments:
I think you have an interesting technique here, but you need to clarify the method, be more honest in the results and data analysis and present what happened more clearly.
Specific comments and suggestions appear above in the various sections.

Additional Questions:
Relevance to the Journal’s Mission: Seems Appropriate to journal

Title:
The title should be changed and the term Traumatic Brain Injury (TBI) removed as only healthy subject are tested.

Abstract:
I do not think the main result as stated is accurate or supported by the data presented in the article. this is a major problem! eg, see fig.4. Native task has superior performance. Discussed more below.

Article Content
The authors need to address the difference between Traumatic Brain Injury patients (TBI) and Mild traumatic brain injury patients (mTBI). Both these terms are used throughout the paper and the patient populations are totally different. These tests appear designed for mTBI pts, which is fine, but make this clear.

The authors are emphasizing motor skill rehabilitation, but the tasks are really perceptual and cognitive in nature - the motor component is high
A major problem from a motor learning perspective is that the motor component of the VR simulation task does not mimic the real world('native') task. Eg, real world task involves grasp and manipulation of small cubes, using dexterous mvts of thumb and fingers. the VR simulation, as currently set up, involves use of a haptic phantom stylus. so the grasp is static, and wrist mvts control the cube position. this is very different from real world movement, and so is a poor model for 'motor' training. The feedback is only 3 dof, missing the torque feedback, making the mvt control & feedback significantly different from the normal healthy situation. However, the cognitive and perceptual - motor learning could be significant - this should be addressed more completely.

Specific comments
on p. 4 authors para 1, describe prior publication on this work, (Li et al 2010) and the difference from present manuscript is not clear. (they sound very similar).
on p. 4, para 2 near end, authors state that users tend to rely on virtual 'fixtures' as a crutch, and then motor learning is impaired - please provide a reference for this statement. on the following page 3, para 1, the opposite viewpoint is presented and referenced - ie, that haptic guidance can enhance learning. If you want to present this as a conflict which your experiemnt is designed to resolve, fine, but you must make this much more clear and provide adequate refs.

Methods/Materials
In general, your figures need legends and more explanation.
apparatus, p. 4 and motor skills test p. 5
The description of the BD test seems pretty clear, but I was unclear on teh ROCF description. in particular, did the subjects only do VR version of ROCF? if so, exactly how were errors calculated? the best design would be to give the ROCF and a different BD subtest in real world setting both before and after the VR and real world training. then the ROCF and BD test could serve as a measure of learning and generalization to a new task - which is what you want to achieve with your training if I understand your purpose correctly. It seems like you mixed baseline testing and training in the first session - but maybe this does not matter if the training was the same as the testing, this is not clear. 3 hours is also a short amount of training for such a complex task. Was Random assignment to experiment groups used? Not clear. The description of the intervention and testing is unclear. eg, on p. 10 i cannot figure out what exactly happened ie each 1 hr training session. how many trials of each task were practiced by each group? hwo many repetitions of each movement were done? How often was feedback provided? were both KP (knowledge of performance) and KR (knowledge of results) used? at what frequency?
The VR features developed by the authors are very nice! As a clinician, and VR user, I think they have the potential to be very useful.

Results
I have a problem with results - presentation and analyses.

Figure 4 seems to be the main result graph. However, I am confused as they show 8 'trials'. is this data for one session only? There were 4 sessions - were there 2 blocks of trials in each session??? task completion time seems > ~10 min at start for VR conditions - is this correct?

Clearly, the native task performance is better overall, yet authors claim that the haptic feedback conditions are superior. The 'rate' of 'learning' is better, but this just represents learning to use the phantom (which would be 10 x harder for patients, by the way, if even able to do so). In addition, the way the sessions were constructed, the VR was actually assisting the subjects with the task, so of course the performance would improve!

On p. 14 authors report repeated measures ANOVA results that essentially show what i say above, yet, they interpret the results as somehow the VR groups being better. they are measuring acquisition here, not learning. (but must clarify the content of fig. 4)
The contrasts used for the pairwise comparisons should be identified. Raw time scores should be included along with the p values - ie, if significant diff in time, say how much. it may be statistically significant, but not clinically meaningful.

The description of the BD test seems pretty clear, but I was unclear on teh ROCF description. in particular, did the subjects only do VR version of ROCF? if so, exactly how were errors calculated? the best design would be to give the ROCF and a different BD subtest in real world setting both before and after the VR and real world training. then the ROCF and BD test could serve as a measure of learning and generalization to a new task - which is what you want to achieve with your training if I understand your purpose correctly. It seems like you mixed baseline testing and training in the first session - but maybe this does not matter if the training was the same as the testing, this is not clear. 3 hours is also a short amount of training for such a complex task. Was Random assignment to experiment groups used? Not clear. The description of the intervention and testing is unclear. eg, on p. 10 i cannot figure out what exactly happened ie each 1 hr training session. how many trials of each task were practiced by each group? hwo many repetitions of each movement were done? How often was feedback provided? were both KP (knowledge of performance) and KR (knowledge of results) used? at what frequency?
The VR features developed by the authors are very nice! As a clinician, and VR user, I think they have the potential to be very useful.

Discussion
You focus on motor learning 'rate' in your discussion but i think what you are really training is perceptual. Well, there is motor control of the phantom, but this is totally unrelated to the motor control a subject would use to do the real task, so its not the sort of motor learning you really need.
The 'rate' is strongly influenced by your VR feedback, which essentially 'pulled' the block to the correct location if the object was close. so, then the score was higher. It is possible I am interpreting this incorrectly but since the description of fig 4 was unclear, I am left to my own devices to interpret.

Basically I don't trust your analysis - it seems like you turned yourself inside out to find a 'significant' result instead of carefully looking at the data and what it might really mean.
I think you have an interesting technique here, but you need to clarify the method, be more honest in the results and data analysis and present what happened more clearly.

Reviewer: 2
Recommendation: Reject

Comments:
The line of research is of interest to wide readership and should be pursued, but the fundamental questions raised in my review point to the very preliminary nature of the described work. Please look more closely at the existing literature in this field and when your experimental design is more solidly justified and issues with experiment design and data analysis are addressed, then this will be an interested study that can further add to the literature on haptic enhanced training for motor skills and rehabilitation.

The paper is well written and clearly described, but I have concerns with the technical approach and methods.

Additional Questions:
Relevancy to the Journal's Mission: The paper describes the early development of a Virtual Reality task environment for training/rehabilitation after TBI. This seems fairly relevant to the journal's mission, although the current manuscript contains no data with the target population.

Article Content: The introduction of the paper describes task analysis as a key fundamental step in developing a haptic simulation for rehabilitation applications. However, the authors don't seem to employ these techniques later in the paper.

The authors fail to adequately motivate the application of robotic rehabilitation to TBI - how is this impairment related to stroke, and why should one assume that techniques and methods from stroke rehabilitation should transfer (or even be investigated) for TBI?

The authors state that the literature is void of information on the relevance of robotic rehab for TBI or how simulation requirements might differ relative to user condition. However, they make minimal contributions towards this end since they only test healthy individuals. Rather, the manuscript describes a proof of concept for task environment and intervention development, not contributing much new knowledge towards the application to TBI.

The authors cite numerous examples of haptic control schemes that can be implemented to accelerate motor learning, such as record and replay, error amplification, and combined approaches. Some key references are missing and in other cases, better references (more accessible to the readership) should be chosen. Most notably the authors fail to include the negative outcomes for haptic augmentation for training. For example, Gillespie has some work on record/replay (virtual teacher) methods that failed to show improvements in motor learning. Li 2007 is a PhD dissertation and should be replaced with more relevant and accessible publications from that group (e.g. O'Malley et al ASME JDSMC 2006; Li et al ACM Trans on Applied Perception 2009), again that show some negative results. There are also more recent references (showing both positive and negative outcomes for haptic augmentation during training) such as Lee and Choi from IEEE Haptics Symposium 2010 who used disturbance, and Huegel and O'Malley showing performance-based haptic guidance (same conference). Later in this section they reference Basdogan, who also has much more recent results (see IEEE World Haptics Conference 2011) with some positive outcomes.

The authors motivate the use of the 'blended' approach to haptic feedback but fail to adequately justify this approach, since other methods have also been shown to be effective (like error augmentation, see recent work by Patton's group, e.g. ICORR 2011).

Methods/Materials There are numerous questions raised about the methods employed by the authors.

First, the choice of the PHANTOM Omni could severely limit the fidelity of the haptic cues in the virtual environment. Therefore, the 'learning' noted in the haptic environments could be due more to the users 'adjusting' and getting used to the low fidelity haptics than actual 'learning' in the motor task. For example, the significant differences in task performance between the native and VR tasks points to limited fidelity of the VR tasks. The authors themselves point to limitations of the hardware specifically the difficulty in achieving desired orientations due to the stylus when manipulating the virtual blocks. I question the selection of this hardware when the goal is to have the VR task performance match native task environment - limited workspace, limited force feedback fidelity, etc compromise this goal significantly.

Define BD before first use of acronym.
some repetition in Apparatus section, likes 37-38.

the use of the 3D immersive environment could also be skewing results - can you justify the use of this technology esp when goal is for implementation with impaired (TBI population) who may have cognitive deficits as well as motor? I think the technology here could be confounding the 'motor learning' and rehab that are the goals.

Did the subjects have any experience with VR or just the training/orientation during the experiment?

were the VR and augmented VR groups equalized at the beginning of the experiment? it seems that they start with one group better than the other - this could be attributed to individual differences rather than a difference due to hardware. I assume that your subject condition is within - in other words, subjects were assigned to only one condition (native, VR, aug VR) - so you need to make sure they are balanced at least the VR vs aug VR groups.

The authors mention that participants were given up to 3 minutes to complete each BD design - how many failed to complete and how was data analyzed when the trial timed out?
In the VR task was the goal to achieve as close to the native environment as possible? If so, say explicitly. You seem to hint at this - make it clearer.

It seems that the visual cues in the aug VR case could have overly constrained the task (certain target locations in the workspace) compared to the other conditions.

The visual cues intended to aid seem potentially complex, again for the target population of TBI this could be confounding results if too much cognitive load when focusing on motor rehab.

Using only % improvement as a metric can skew your findings. If the VR tasks are significantly limited in terms of fidelity (which I expect they are, given hardware and comments above) then the % improvement could be training to deal with the limited fidelity, not just motor learning.

Were the pre-post evaluations done in the same environment as training? All that would show is that people training in VR get better in VR. Why compare to Native? Limited fidelity hardware will mean that performance will likely never achieve native performance, so this is unfair comparison.

How were subjects assigned to conditions? Were they randomized?

Why a linear model fit to the learning trends? Much literature indicates power curves for learning, not linear fits. Even your plotted trends look power/exponential not linear. I question the value of this analysis at a very fundamental level. How good were these fits? Why compare intercepts? All that shows is that initially the groups were different, which is to be expected since VR does not equal native. And you show that your aug VR is not that different from VR. Not surprising. Comparing slopes again is not justified since learning curves are power/exponential.

Discussion Not much to say about discussion - I have serious concerns and reservations about the methods of experimentation and analysis, so comments on discussion are not valuable until those base concerns are addressed.

Editor's Comments to Author: The title of the manuscript is misleading. No TBI subjects were tested. The motor skill training was not an accurate description of the study tasks which had significant perceptual and cognitive components. The rationale of the study needs to be better justified. It was not clear how and why the VR+haptic training was particularly suitable for people with mTBI. As pointed out by the two reviewers, the methodology and analysis sections missed details and were difficult to follow. The study protocol and subject assignment procedure should be clearly described. Instead of the change scores, the baseline and post-test scores should be provided. In general, if the authors consider resubmitting the manuscript, some underlying rationales and purpose of conducting this study should be explicitly and well justified and explained. The methodology and statistical analysis also need significant improvement in terms of clarity and scientific rigor.