Reviewer #1: Major Comment:  
While they show significant differences, the differences are miniscule (e.g. Fig. 2, D & E, and in most other figures). These miniscule differences are significant because they are based on many thousands of data points.  
When Computers of Average Transients (CAT) were introduced to neuroscience ~60 years ago, somebody measured the average response of the foot-toe to tens of thousands of stimuli and found significant (but miniscule) response. Does that means the toe is involved in perception of the stimulus?  
The same is true here. The results show that WM task affect the hippocampal activity, but does not "demonstrate its involvement" in the task. This situation should be explicitely stated in the discussion. All such statements in the abstract, summary, introduction and discussion should be changed to something like: "This study suggests that …"  
Other Comments:  
1. The methods with which they processed the data are extremely hard to follow. Although they provide access to their software I suspect most reader would not go through the trouble of studying their software. I do not know what can be done about that. Perhaps the editor can allow extended Supplementary section in which they can explain and illustrate what each type of processing do to the data.  
2. The style of the results makes it very difficult to follow due to the heavy mixture of statement with statistical parameter. For example (lines 175-178]: The trajectory distance from the origin (O) of the hippocampus (1.11 [1.01]; median [IQR]; n=195,862 time points) was larger than those of the EC (0.94 [1.01]; median [IQR]; n = 165,281 time points (figure C & D).  
Perhaps putting the statistic data as a comment at the bottom of the page, or using a different font to the statistic can help.  
3. It is not clear how the multi-unit activity was binned. Are the 50 ms bins non-overlapping, or they are partially overlapping? In the first case results may seriously depend where the border between bins is placed.. in the latter the "n time points" are not independent.  
  
Minor comments:  
Figure 4: There are No A, B, … in the fugure.  
Figure 5 A: I do not see +/- 95% limit  
Figure 6, lines 603,604: I do not see the madian on the X axis.  
  
  
  
Reviewer #2: While the authors do provide some background on our current understanding of the role of the hippocampus in some kinds of working memory, they do not use that background to motivate the hypothesis or techniques presented in the article. It is not clear what new understanding might be gained by proving or disproving the hypothesis, and many of the methodological choices are simply stated without any discussion of why they are (or might not be) the \*right\* choices.  
  
A couple of high-level questions it would have been nice to see addressed in the text:  
  
1) What (simpler) analyses have been performed on this dataset already? More importantly, given the outcomes of those analyses is there any chance that the hypothesis put forward in this article could be wrong?  
  
2) Why is GPFA an appropriate dimensionality reduction technique for this dataset, what alternatives were considered (and why were they rejected), and what alternatives were actually tried (and what were the results)?  
  
Further, the text as a whole could use some revision with an eye towards clarity of meaning. That service will not be provided as part of this reviewer's notes. But as an example, the ending of this (rather important) sentence is superfluous and confusing: "We hypothesized that multiunit activities in the hippocampus alter their representations depending on the memory load and phase of a WM task, particularly in relation to sharp-wave ripple complexes (SWRs), which are known to be various cognitive biomarkers."