Reviewers' comments:  
  
Reviewer #1: This is a new version of the  manuscript by Karimi-Rouzbahani et al, about the neural encoding of facial familiarity using EEG and MVPA, which i have already reviewed before to another journal.  
Authors worked on the mscript and addressed some of my original questions. Here i just list the ones i still find unaddressed or which i discovered now.  
  
I already found the the article interesting, worth of publishing, clear and using solid methods for the first time and i think this version is definitely improved and would certainly be important to see published.  
  
Still, i keep my original major comment important and although in a hidden form it is addressed in the manuscript now, i wish it would be more detailed and commented on.  
  
Major comment  
My only major problem with the results lays in the simple interpretation of anterior contributions to the encoding of familiarity as feed-back. You find, using a clever partialling out method, that eliminating the occipital contributions from the frontal (or rather anterior, as it involves temporal cortex too) electrode pattern familiarity decoding reduces stronger and earlier-longer information encoding about familiarity, when compared to the opposite, when you partial out the frontal information from that of the occipital/posterior electrode pattern. The former is interpreted as a signal of feed-back, while the opposite as feed-forward information flow. This makes sense but only if the frontal cortex does not play a role, on its own right, in face processing. However, the inferior frontal face area (see e.g. Collins and Olson,2014) is known to be associated to the STS and playing a role in social, dynamic and eye-movement related information processing. If we assume that these tasks are more related to the frontal than to the posterior areas, as for example Duchaine and Yovel, 2015 do, then the results of the partialling out analysis merely mean that the functions of the frontal areas are modulated more by the posterior areas (in other words, in those functions the parietal areas also play a role) than the other way around. The lower-level functions of the posterior sites are, on the other hand, modulated less, shorter, later by the removal of frontal areas, in other words the frontal cortexes do not play much role in them.  
This is different from your conclusion where you state feed-forward vs feed-back connections.  
I dont see any good way to come around this alternative (and simpler) conclusion than your assumption about connectivity.  Time would be a potential factor to resolve it, feed-back being later, but in your figures it is clear that the two periods overlap entirely and the peaks are also almost fall into identical windows. Unless i overlooked something and you can give a convincing way to exclude this possibility i would recommend to a) discuss this in the paper and b) tune down your respective conclusions throughout the manuscript.  
  
Minor  
  
Fig 1 b and c lacks standard errors.  
Fig 1 D: why do the famous faces start at around 30% correct performance when 22 % coherence is given? This actually means that these faces are 70% misjudged as unfamiliar. In other words, you have a strong bias towards unfamiliar responses at this condition. Why? Could this affect your results?  
I find it interesting that you find no difference between self and personally familiar in behaviour. In fig 2b it is clear that the erps are gradual, on the other hand. I would ask you to discuss this issue briefly.  
Fig2. why not presenting all the conditions you have separately? Same for fig 3. I mean why merging the conditions? You have a 2 x2 design essentially with coherence and familiarity. I think it would be more useful for the readers to see all the conditions.  
Your fig 4a is not too easy to understand. The RDMs are fine, but the right side is not too informative.  
  
  
  
  
  
Reviewer #2: I read the paper "Perceptual difficulty modulates the direction of information flow in familiar face recognition" with interest.  
I think the topic is interesting and the authors have used different analysis techniques to explore the question.  
I have a few comments that I list here, the order is from intro to references of the paper:  
1)      I like the terms "peri-frontal" and "peri-occipital" which correspond to groups of electrodes. I searched for the term and it seems like only these authors have used it in the past. Therefore, I think the term needs to be explained explicitly in a footnote or in parentheses the first time that it is mentioned (line 43). Also is this rather gross grouping common? if yes, please provide the references.  
2)      Line 67: represented better or worse: I find that a bit sloppy for an article. Better/worse is too vague and could be replaced with other words.  
3)      The design is very rich, in the sense that both familiarity and coherence are somehow parametrically varied. Nonetheless, the authors have always pooled along one dimension, possibly obscuring any interaction effect. For example, Figure 1 (B-E) could be replaced with matrices that show as images, behaviour vs. both factors.  
4)      I tried the demo link (line 207) and it didn't work.  
5)      The ERP analysis are interesting, but they only show the results for CP2. Topoplots could be shown (in the supplementary) to give an idea about the localisation of the effect.  
6)      Are the CP2 effects driven by decision uncertainty (similar to what others have reported about CPP)? Ruling this out would be important for interpreation.  
7)      Once again, Figure 3 could all be matrices. However, the current format could also be improved if they added non-parametric inference on clusters (to find significant clusters in time).  
8)      About the information-flow analysis:  
what the authors depict in Figure4 is that for FF they get the partial correlation between peri-frontal at time T and peri-occipital at time T-delay, after partialling out a model. What is that model? Why did they choose that model? How do they determine the delay?  
9)      The same questions also apply to Figure5. Additionally, I have extra concerns about the models used in Figure 5 (especially the last three from left to right). The models make comparisons between very unbalanced number of dissimilarity pairs. I have the worry that difference in SNR could drive the correlations. This would not have been an issue if they used cross-validated squared Euclidean distances to construct RDMs (since they are sensitive to variance differences).  
10)     You cite our recent paper with Alessio and others. Though the title starts with "Multi-dimensional connectivity".  
11)     I'm not sure if Eq 2 and 3 follow the standard math notations.  
12)     Although I agree with the authors that the previous methods were not perfectly suited for their information-flow analysis, I think a Granger RSA approach (like Kietzman etal. PNAS) could have also been a possible option. The simplest would be to correlate your ROI RDMs with the model RDM and do a Granger Causality on the time courses. If this is not equivalent to what you are doing in your partial-correlations, then it'd be best to spell it out more clearly, so that others get a deeper insight into your novel method and the paper also more clearly provides a novel method.