Author's response to reviews

Title: 3D facial landmarks: Inter-operator variability of manual annotation

Authors:

Jens Fagertun JF (jenf@dtu.dk)
Stine Harder SH (sthar@dtu.dk)
Anders Rosengren AR (anders.rosengren@regionh.dk)
Christian Moeller CM (christianm9@gmail.com)
Thomas Werge TW (thomas.werge@regionh.dk)
Rasmus R Paulsen RRP (rapa@dtu.dk)
Thomas F Hansen TFH (thomas.hansen@regionh.dk)

Version: 3
Date: 21 August 2014

Author's response to reviews: see over
Dear Dr. Sedda

Please find enclosed the revised manuscript, *3D facial landmarks: Inter-operator variability of manual annotation*.

We have addressed the comments point-by-point below. The revised manuscript has subsequently been proof-read by an native English speaker with scientific experience. We hope you find the revised manuscript worthy of publication in your journal.

Yours sincerely

Thomas Hansen
Institute of Biological Psychiatry, Mental Health Centre, Sct. Hans
Copenhagen University Hospital
Boserupvej 2, DK-4000 Roskilde
Comments to Author (Italic).

Page 2 line 4: it might be useful to explicitly state also in the introduction what are the risks of having too much variability when annotating facial landmarks. I assume the studies in other fields approached the problem because of clinical/practical reasons. The authors should stress this point also regarding their study.

We agree and in the revised manuscript, we now state the importance of understanding the variability introduced by manual annotation. See Page 2 Line 2 and Page 7 Line 1.

Page 2 line 28 and later: it is not clear how 36 facial scans were selected from the starting 41. Why were some scans excluded? What random allocation method has been used?

We apologies for lack of information, and have now included this information in the revised manuscript. See Page 2 Line 29.

What was the mean age of the persons depicted in the scan? (facial morphology changes with age and could affect variability in annotation). The authors should provide these important details. Further, this seems a small dataset of stimuli.

We agree that the age could have an effect on the reliability and gender as well. We did consider including age and gender as individual variables. However, given the size of the dataset we chose to include portrait number instead, which should capture any effect as to the picture itself, including the putative effect e.g. age and gender. We have elaborated on this in the discussion in the revised manuscript, see Page 6 Line 36.

Page 3 line 38: “pilot analysis...data not shown”. As data are not shown and very few information is provided, the authors might consider dropping this sentence.

We have deleted this sentence in the revised manuscript.

More information about the operators should be provided, including their experience (years of working might be a good measure), whether they were blind to the aim of the study, etc. In particular experience should be introduced as a covariate in the model to further understand the influence of inter-operator variability.

We apologies for not making this clear. This has been clarified in the manuscript, both in the methods and results see Page 3 Line 32 and Page 4 Line 37.

The authors of the paper have contributed to different steps of the development of the current manual annotation tool and standard. We therefore chose a common training program, prior to the annotation. Thus, we unable to estimate the effect of experience given these premises.

Page 5: the authors only report the standard deviations of their ANOVA model and for the model fitting. I am a bit confused by the absence of probability values. Is there a reason why these are not reported?

We apologies for the confusion, we have improved this in the revised manuscript. The p-values are not provided as the aim of the study is to identify variability for use in future studies; in effort to minimized the variance in these and to estimate noise-to-signal ratio, regardless of significance.
Further, what statistical software did the authors use?  
We apologies for lack of information, we have now included this in the revised manuscript, see Page 4 Line 29.

While the adoption of only Caucasian ethnicity might be a limitation, it should be also noticed that is it is an advantage in terms of homogeneity of the experimental stimuli used (minimizing the role of stimuli variance).  
We appreciate the comment and we agree that one would expect the ethical homogeneity to minimize the stimuli variance. However, we chose not to comment on this, as we cannot excluded that there might be other levels stimuli within the pictures that are not captured within ethnicity.