

Exploring the potential of analysing visual search behaviour data using FROC (free-response receiver operating characteristic) method: an initial study

Leng Dong^a, Yan Chen^{*a}, Sarah Dias^b, William Stone^b, Joseph Dias^c, John Rout^b, Alastair G. Gale^a

^aApplied Vision Research Centre, Loughborough University, Loughborough, UK;

^bUniversity of Birmingham, UK

^cUniversity of Leicester, UK

ABSTRACT

Visual search techniques and FROC analysis have been widely used in radiology to understand medical image perceptual behaviour and diagnostic performance. The potential of exploiting the advantages of both methodologies is of great interest to medical researchers. In this study, eye tracking data of eight dental practitioners was investigated. The visual search measures and their analyses are considered here. Each participant interpreted 20 dental radiographs which were chosen by an expert dental radiologist. Various eye movement measurements were obtained based on image area of interest (AOI) information. FROC analysis was then carried out by using these eye movement measurements as a direct input source. The performance of FROC methods using different input parameters was tested. The results showed that there were significant differences in FROC measures, based on eye movement data, between groups with different experience levels. Namely, the area under the curve (AUC) score evidenced higher values for experienced group for the measurements of fixation and dwell time. Also, positive correlations were found for AUC scores between the eye movement data conducted FROC and rating based FROC. FROC analysis using eye movement measurements as input variables can act as a potential performance indicator to deliver assessment in medical imaging interpretation and assess training procedures. Visual search data analyses lead to new ways of combining eye movement data and FROC methods to provide an alternative dimension to assess performance and visual search behaviour in the area of medical imaging perceptual tasks.

Keywords: performance, eye tracking, area of interest, FROC

1. INTRODUCTION

Radiology plays an important role in medical imaging diagnosis. Usually the responses of radiologists are acquired as they inspect a series of cases and then these data are subsequently analysed to tease out particular performance variables of interest to the investigation. Additionally, eye tracking methods have been widely used in radiology to understand medical image perceptual behaviour and examine why interpretation errors occur. Various eye movement parameters, such as dwell time and number of eye fixations, can be calculated together with an area of interest around a potential abnormality to describe whether an observer's visual attention has been located correctly on suspicious areas. Previous studies have suggested that, for instance, in the area of mammography reporting, some 70% of breast cancers were in fact visually inspected by radiologists¹ but still missed. However, current methods of eye-tracking analyses mostly focus on the actual values of eye movement measures which are not threshold-bias independent².

Typically, perceptual studies in radiology utilise a series of cases, some of which are abnormal and some of which are normal, often the case mix being 50% of each. The current work explores whether eye movement data alone can be demonstrated to be useful in assessing radiological performance. To do this we used a dataset where a number of observers of different skill levels had visually searched a series of images and where their eye movements were tracked and they also gave rating response information about each image after inspecting it. These particular dataset cases all contained one or more abnormalities thus rendering the typical ROC analytic approach inappropriate.

* y.chen@lboro.ac.uk

Free-response receiver operating characteristic (FROC) measures of observer performance are normally provided by collecting ratings from the observer as well as the perceived suspicious regions. Those marked regions can be classified as lesion-localizations (LLs) or non-lesion localization (NLs) by using lesion localization information provided either from an expert or from the case biopsy results as a gold standard measure.

The Area under the ROC Curve (AUC) can be worked out as a figure-of-merit (FOM) which eliminates threshold-bias by measuring the differential ability of the ratings to correctly classify diseased and non-diseased images³. However, the ratings collected are decided by the observer subjectively which may not be considered fully representative of an observer's perceptual behaviour. Also, the rating scale suggested in some circumstances (e.g. BIRADS in mammography reporting) is not considered as an ideal rating system for ROC/FROC analysis as the diagnostic confidence decision class fundamentally requires to be reported on an ordinal scale⁴.

Based on the above considerations, it is proposed that by using the FROC methodology to analyse eye tracking data, the advantage of both methodologies can potentially be exploited. A similar study² involving eight expert breast radiologists interpreting mammograms suggests that there is agreement between rating and eye tracking data in FROC analysis.

In the current study, the data collected involved eight health professionals of different experience levels reading a case set of 20 radiographs. The study investigated whether the utilization of eye-tracking techniques and FROC methods alone, could help identify differences in skill. Such FROC measures were also compared to observer response based FROC analysis.

2. METHOD

Twenty medical images first were selected by an expert. Each case contained one or more potential abnormal areas which were delineated by the expert. For the current purposes data from eight participants, consisting of four experienced and four less experienced health professionals were employed. Each participant had examined the series of cases whilst their visual search behavior was monitored and for every case they made clinical decisions. The radiographs were viewed on an LCD monitor. During each trial the participant's eye movements were tracked by a Tobii X50 eye tracker placed underneath the monitor.

Three different eye movement measurements were extracted and clustered using ClearView software. The data obtained were classified into fixations, dwells and clusters, defined below.

- Fixations: The number of fixation points within the area of interest.
- Dwells: Time spent fixating within the area of interest around an abnormality site.
- Clusters: The number of continuous fixation points within the AOI

FROC analysis was then applied to the above eye movement measurements with the AOI providing location information to determine whether any agreements or differences could be found between the two groups. The values for those measurements inside the AOI are considered as Lesion Localizations (LLs) rating and correspondingly the same measurements outside AOI are used as Non-Lesion Localizations (NLs) ratings. The values of the area under the curve (AUC) were computed offering a measure of performance accuracy that considers both sensitivity and specificity. The values of AUC here range from 0 to 1 as the LLs ratings are not necessarily higher than the NLs rating. This is different from the conventional ROC curve which has an AUC measure ranging from 0.5 to 1.

Independent samples t-test was then used to measure the significance of any differences between the two groups. Also mark-rating data based FROC analysis was carried out by using severity rating and location information. Severity ratings on those cases where the participant successfully spotted the lesion area are treated as LL and severity ratings on the case where participants missed the lesion area are treated as NL. Pearson correlation was finally conducted to test the relationship between eye tracking FROC and rating based FROC results.

Figure 1 demonstrates the procedure of data collection and analysis for this study. Participants examined the images, responding with abnormality rating and location information, which yield physical LL and NL measures which are then used in the usual FROC analytic approach. Whilst examining the images the participants' visual search behaviour is monitored and data subsequently extracted concerning fixations, dwells and clusters with regard to known abnormality locations which then yield virtual LL and NL marks which are then used in FROC analyses. Correlation analysis was then used to compare the results from both approaches.

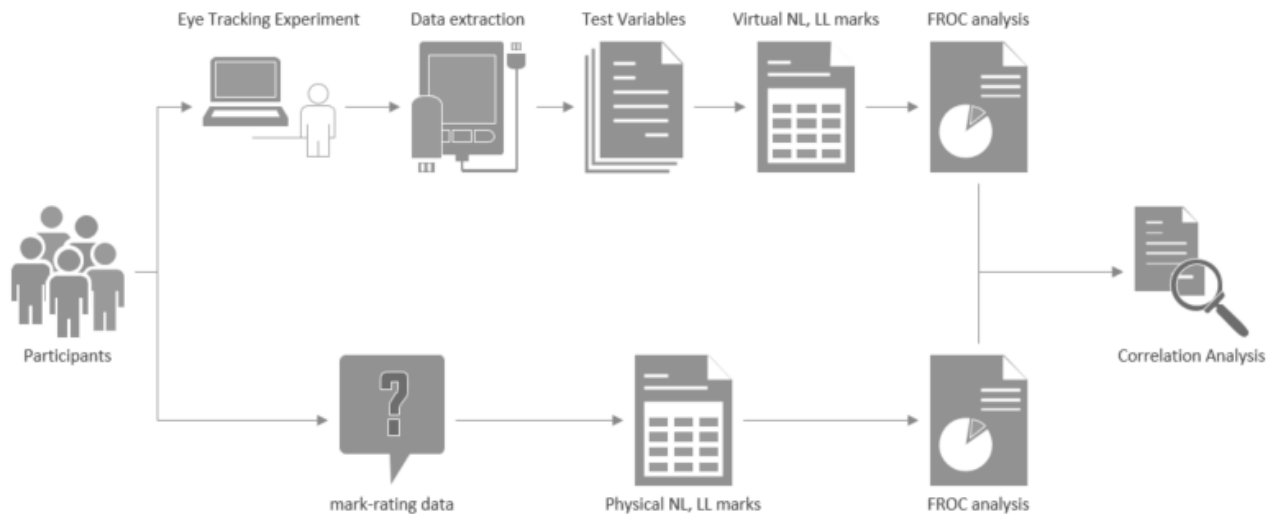


Figure 1 Diagram of the data collection and processing method to conduct FROC using eye movement measurements as input variables.

3. RESULTS

FROC analysis using eye tracking measurements as virtual marks implicates that, for all three measurements shown in Figure 2, the inexperienced group scored the lower average values of AUC (Fixation: $M=0.3375$, $SD=0.0580$; Dwell: $M=0.3675$, $SD=0.0781$; Cluster: $M=0.5025$, $SD=0.0206$) than the experienced groups (Fixation: $M=0.5350$, $SD=0.0370$; Dwell: $M=0.5775$, $SD=0.0591$; Cluster: $M=0.5250$, $SD=0.0100$). For measurements of fixation and dwell data used as FROC virtual ratings, independent samples t-test shows strong statistical differences ($p<0.01$) of average AUC between two groups (fixation: $t(6)=-5.746$, $p=0.001$; dwell: $t(6)=-4.290$, $p=0.005$). For measurement of cluster data as virtual ratings conducted FROC, no significant difference ($p>=0.05$) of average AUC was spotted according to the independent samples t-test (cluster: $t(4.338)=-1.964$, $p=0.115$).

Mark-rating based FROC analysis was conducted by using rating provided by participants subjectively and location information which categorized rating into LLs and NLs. As shown by the box plot in Figure 2, the AUCs calculated indicate that the inexperienced group ($M=0.6575$, $SD=0.0789$) have the lower average AUC than experienced group ($M=0.9575$, $SD=0.0378$). There was a strong significant difference ($p<0.01$) between two groups as determined by independent samples t-test ($t(6)=-6.860$, $p<0.001$).

A series of Pearson product-moment correlations were run to determine the relationship between rating rating-based FROC and eye-tracking measurement conducted FROC. Statistically significant positive correlations were found for eye movement measurements of fixation and dwell based FROC compared with rating based FROC (rating vs fixation: $r=0.817$, $p=0.013$; rating vs dwell: $r=0.747$, $p=0.033$). Increases in AUC of fixation and dwell data conducted FROC were correlated with increases in AUC of rating based FROC. Correlation between cluster data conducted FROC and rating-based FROC was not significant ($r=0.540$, $p=0.167$). The relationships between AUCs of rating based FROC and eye movement measurements conducted FROC are shown as scatterplots in Figure 3 with regression lines and the relevant equation also given.

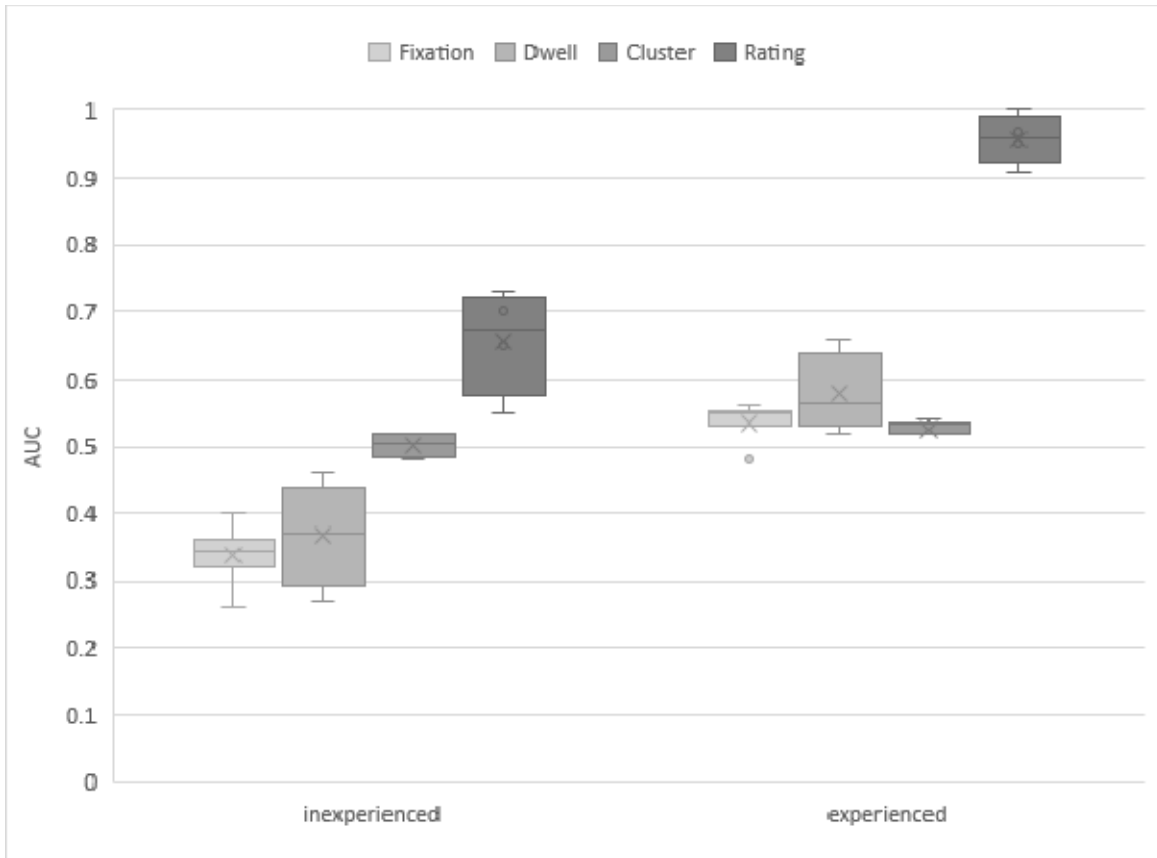


Figure 2 Box plots of AUCs from both eye movement data conducted FROC and rating-based FROC comparing inexperienced and experienced groups

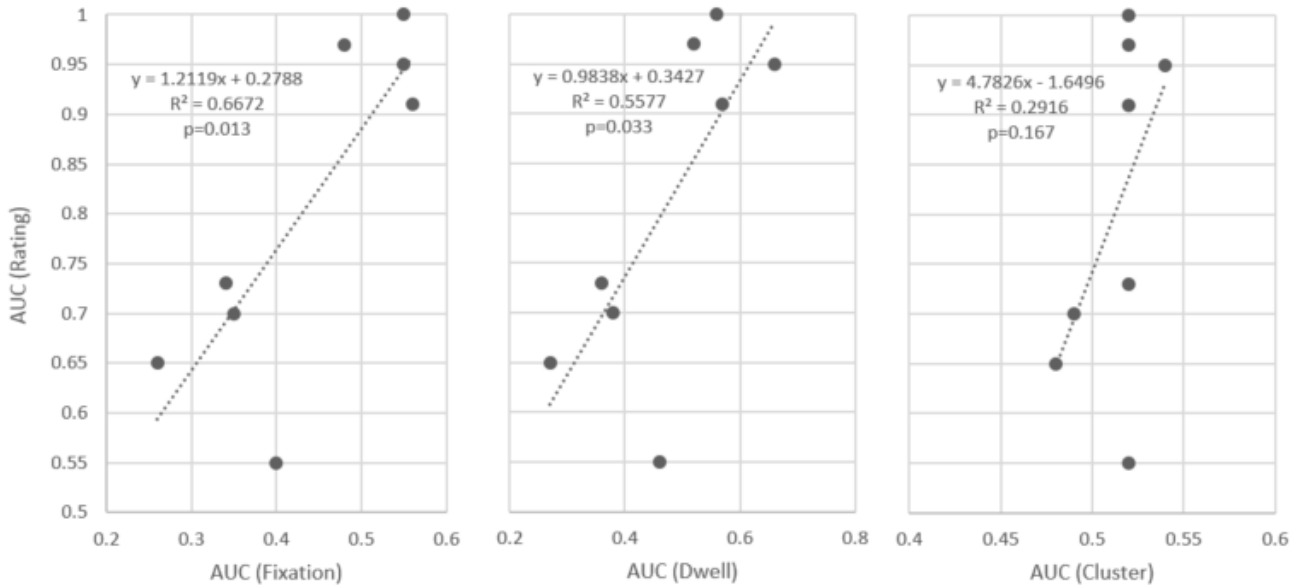


Figure 3 Scatterplots and regression lines indicating relationship between AUCs of rating based FROC and eye movement measurements conducted FROC

Additionally, a paired-samples t-test was conducted to compare eye movement measurements conducted FROC and rating based FROC. There were significant differences in each pair of AUC scores for rating based FROC (M=0.8075, SD=0.1703) and eye movement measurements conducted FROC (fixation: M=0.4363, SD=0.1148, $t(7)=10.376$, $p<0.001$; dwell: M=0.4725, SD=0.1293, $t(7)=8.366$, $p<0.001$; cluster: M=0.5137, SD=0.0192, $t(7)=5.170$, $p=0.001$). These results suggest that AUC scores for rating-based FROC are significantly higher than eye movement measurements conducted FROC tested in this study. Table 1 lists the summary of average values of AUC scores for each test variable together with standard deviation.

Table 1 summary of average values of performance scores for each experience groups together with standard deviation

Group	Fixation (AUC)	Dwell (AUC)	Cluster (AUC)	Rating (AUC)
Inexperienced	0.3375±0.0580	0.3675±0.0781	0.5025±0.0206	0.6575±0.0789
Experienced	0.5350±0.0370	0.5775±0.0591	0.5250±0.0100	0.9575±0.0378
Total	0.4363±0.1148	0.4725±0.1293	0.5137 ±0.0192	0.8075±0.1703

4. DISCUSSION

A methodology is presented here based on partial data from a larger study in which participants of differing experience had examined a series of medical images. The cases had all contained abnormalities which then rendered typical ROC analyses impossible as all cases contained a potential target. Whilst sensitivity could be assessed it was not possible to assess specificity. Also, potentially all the observers had to do was to respond ‘abnormal’ for every case in order to be correct as scored by ROC analysis. FROC analysis overcomes this for these data as it also considers location information of observer responses as compared to known AOIs around abnormality sites. Using eye movement data alone demonstrated that differences between the two groups of experienced observers could be demonstrated. There were significant differences between the AUC scores for the ratings based analyses and the eye movement based AUC. However, as shown in Figure 3, two out of three AUC eye movement measures here were correlated with the ratings AUC measures thereby indicating that eye movement data alone could be used in FROC analyses of observer studies’ data although doing so will yield lower AUC measures than using ratings alone.

5. CONCLUSIONS

This initial study demonstrates that applying FROC analysis using eye movement data as input variables can act as a potential performance indicator in medical imaging studies. A significant difference was found between the inexperienced group and the experienced group when using fixation and dwell time as virtual marks conducting FROC analysis. Additionally, agreements were found between the eye movement measurements conducted FROC and rating-based FROC when using fixation and dwell data as input variables. This then interestingly suggests that visual search parameters alone can potentially be employed instead of response based measures to assess skilled performance.

REFERENCES

- [1] Kundel, H., Nodine, C., & Krupinski, E. (1990). Computer-Displayed Eye Position as a Visual Aid to Pulmonary Nodule Interpretation. *Investigative Radiology*, 25:890-6.
- [2] Chakraborty, D., Hong-Jun, Y., & Mello-Thoms, C. (2012). Application of threshold-bias independent analysis to eye-tracking and FROC data. *Academic Radiology*, 19: 1474-83.
- [3] Metz, C. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4): 283-98.
- [4] Jiang, Y., & Metz, C. (2010). BI-RADS Data Should Not Be Used to Estimate ROC Curves in Radiology, 256:29-31