
Downloaded from: http://insight.cumbria.ac.uk/id/eprint/404/

Usage of any items from the University of Cumbria's institutional repository 'Insight' must conform to the following fair usage guidelines.

Any item and its associated metadata held in the University of Cumbria’s institutional repository Insight (unless stated otherwise on the metadata record) may be copied, displayed or performed, and stored in line with the JISC fair dealing guidelines (available here) for educational and not-for-profit activities provided that

- the authors, title and full bibliographic details of the item are cited clearly when any part of the work is referred to verbally or in the written form
  - a hyperlink/URL to the original Insight record of that item is included in any citations of the work
- the content is not changed in any way
- all files required for usage of the item are kept together with the main item file.

You may not

- sell any part of an item
- refer to any part of an item without citation
- amend any item or contextualise it in a way that will impugn the creator’s reputation
- remove or alter the copyright statement on an item.

The full policy can be found here. Alternatively contact the University of Cumbria Repository Editor by emailing insight@cumbria.ac.uk.
Implication of the first decision on visual information-sampling in the Spatial Frequency domain in pulmonary nodule recognition

Mariusz W Pietrzyk*1,2, David Manning1, Tim Donovan1,3 and Alan Dix2

1School of Medical Imaging Sciences, University of Cumbria, Bowerham Road, Lancaster, LA1 3DJ, UK; 2Computing Department, Lancaster University, Lancaster, LA1 4YR, UK, 3Psychology Department, Lancaster University, LA1 4YF, UK

ABSTRACT

Aim: To investigate the impact on visual sampling strategy and pulmonary nodule recognition of image-based properties of background locations in dwelled regions where the first overt decision was made. Background: Recent studies in mammography show that the first overt decision (TP or FP) has an influence on further image reading including the correctness of the following decisions. Furthermore, the correlation between the spatial frequency properties of the local background following decision sites and the first decision correctness has been reported. Methods: Subjects with different radiological experience were eye tracked during detection of pulmonary nodules from PA chest radiographs. Number of outcomes and the overall quality of performance are analysed in terms of the cases where correct or incorrect decisions were made. JAFROC methodology is applied. The spatial frequency properties of selected local backgrounds related to a certain decisions were studied. ANOVA was used to compare the logarithmic values of energy carried by non redundant stationary wavelet packet coefficients. Results: A strong correlation has been found between the number of TP as a first decision and the JAFROC score \( r = 0.74 \). The number of FP as a first decision was found negatively correlated with JAFROC \( r = -0.75 \). Moreover, the differential spatial frequency profiles outcomes depend on the first choice correctness. Keywords: Image Perception, Spatial Frequency Analysis, Wavelet, Eye-tracking

1 INTRODUCTION

Recent findings [1][2] support a paradigm shift within visual perception model in medical imaging underlining the importance of the holistic view advantages the visual searching stage. The perception of medical images starts with global impression. This is very rapid process held in the first hundred milliseconds of onset to the observer. This time is long enough for transforming and processing visual information in the eye-brain system and reaching the visual cortex. However, this period, it is too short to launch high level processes in the brain, which are related to more conscious analysis and interpretation. These high cognitive neurological-algorithms are typical in the following, a longer run phase of perception – discovery scanning. The time scale of the perception in holistic mode supports the view of an automated or spontaneous response during those 100ms. Discovery scanning is characterised by eye movements (saccades) directing gaze to a particular region of interest (ROI) and a longer time dedicated to selected sites (fixations) during which subject attention has a radial limited up to foveal size. This assumption that visual attention is allocated with the prolonged dwelled sites [3] has been a widely accepted paradigm in medical image perception and radiological decision-making studies [4][5]. Thus, eye tracking methodology can contribute to understanding perceptual-cognitive mechanisms behind visual searching strategy and decision-making [6]. The global-focal perception hypothesis focuses on the perception of various types of information perceived at different parts of the time-spatial frame by the observer’s visual field in a particular temporal task. These information-types are well represented in terms of spatial frequency properties by a low spatial frequency (LSF) characterizing the global information about visual stimuli and high spatial frequency (HSF) components related to the details of image. Moreover, what has been shown is the LSF features are processed and analysed much faster by the observer brain (in a few hundreds msec) than HSF, which takes seconds [7]. A model divides the perception into sequential phases (Figure 1.). So our contention is that, spatial frequency properties are crucial for processing and understanding visual information [8].

* m.pietrzyk@lancaster.ac.uk
Figure 1. Graphical representation of the visual searching and detection model proposed for radiological task [9].

During the first few hundred milliseconds a global impression about the image has been created by the brain, often without eye movements and without consciousness of the mechanism. This holistic analysis uses the low spatial frequencies extracted from the entire retinal network, and some suspicious regions are believed to be flagged during the first stage of perception [10]. That flagging mechanism works as an automatic feature detection schedule and forms a mental hypothesis of the perturbation localisations and is based on previous experience, knowledge and skills as well as an intuition and talent. However, the early formation of hypotheses will possibly incur errors in perception. After this very short phase the observer directs the high acuity beam of visual field into the specific locations of previously mapped regions to perform more detailed analysis.

The global impression seems to influence the radiological performance.. The flagging mechanism has been proposed as an attention distribution framework. Further, the very first decision made about the presence of abnormality influences the rest of the radiological performance in that case. Similarly, about two decades ago, Satisfaction of Search (SOS) was recognized. This effect manifests as a significant decrease in the probability of detecting a second abnormality, e.g. second nodule visible at the chest radiograph [11]. That human factor limitation can be handled with double reading [12]. Unfortunately, none of the current CAD systems deal with the SOS effect [13], so a second expert observer is normally required for double reading.

More recently, the effect of the type of the first decision on the overall performance has been found in mammography [2]. The implication of that rapid perception phase has been investigated in terms of the influence on the accuracy in decision-making [14]. Moreover, the effect of a misleading global impression on the first decision results in incorrect marked normal regions. In contrast, the true-positive (TP) response follows a correct holistic interpretation in the first hundreds ms of viewing an image. Results support a conclusion of a consistently accurate course in the decision-making process if only the first decision is TP. In case of false-positive (FP) as first decision, the subject has an average 6% chance to change that unfortunate tendency. Spatial frequency analyses support the hypothesis of the internal template [15] serving as a mental kernel in the convolution on a foveated location, where the result corresponds with the level of confidence in the overt decision scale. Once, the mental kernel is set up with the first decision, the following responses are related to the local spatial frequency features biased in favour of the similarities with the first decision. For that reason, a cascade effect may explain. The mental kernel is believed to be developed through years of experience. If this is the case a more significant cascade effect is expected to be found in experts’ performance compare with less experience subject.

The reports from these experiments present a compelling idea for selective spatial frequency encoding of areas of mammography images that experienced observers find visually salient in the search for pathology. However, what is not yet known is whether this is a general phenomenon in visual-cognitive decision-making in radiology or whether it is something unique to the field of breast imaging cancer detection. Since mammography is a modality with a range of special characteristics compared with mainstream radiological images there is a strong case for investigating the spatial frequency phenomenon and its consequences in other common areas of radiology.

This study focuses on aspects related to the detection of pulmonary nodules from Posterior-Anterior Chest Radiographs performed by observer with different levels of experience. First, the visual sampling during discovering scanning preceding first decision is analysed in terms of the type of the information carried by different spatial frequency bands. The physical properties of those sites – dwelled before first hit – is conducted in terms of the decision’s
correctness and the spatial frequency features from local background. Second, the overall performance measured by JAFROC FOM is analysed in terms of the number of the TP and FP as a first decision. Then, the eye tracking data, which characterise the decision sites (like dwell time, time event or confidence level) are investigated in terms of the type of the first outcome in the case. Finally, the implication of the first decision is analysed on the scale of similarities in the energy distribution in the Fourier plane between different outcomes following a TP or FP.

1.1 RESEARCH AIMS

- Correlation between the JAFROC FOM and the number of TP and FP occurred as a first decision
- To investigate the influence of the first order decision on the similarities between different outcomes (TP, FP, FN) in spatial frequency domain and the related eye tracking data (time event, dwell time, confidence level)

2 METHODOLOGY

The experiment was designed to investigate the subject’s performance during a radiological task of the detection and recognition of the pulmonary nodules from Posterior-Anterior (PA) chest radiographs (CXR). The retrospective analyses of the radiological cases were done by the participants with different levels of experience in the task. The eye tracking device was used to collect the ocular-motor, viewing conditions and the data on overt decision locations of observers.

2.1 EYE TRACKING EXPERIMENT

The eye tracking experiments were conducted involving forty subjects (n = 40) with different radiological experience (Experts, 2 levels of Novice and Naives). Experts represent the most experienced professional observer: 8 radiologists and 2 reporting radiographers: age range 35 – 55y (mean 47.2). Two novice levels were recruited from 3rd and 2nd year of the BSc Radiography Course at University of Cumbria. First students’ group (age range 20-54, mean 29.4) would have 28 weeks clinical experience, whereas second group (age range 19-48, mean 31.6) at least 12 weeks. The naïve observers, came from disciplines not related with radiological tasks (age range 22-40, mean 29.2). Each group contains 10 observers. The stand-alone Tobii x50 eye tracker together with a 19-inch LCD monitor was used in experiments in dedicated laboratory room with fully controlled light conditions set up to 30 lux. This study is based on a test bank of chest radiographs: 20 digitised gray scale (8-bits) images, with 50% of prevalence of films with highly salient or camouflaged nodules which are histopathologically proven (Figure 2).

Figure 2. The pulmonary nodule localizations in the test bank of PA chest radiographs. 10 radiographs contain 23 nodules in total, within 5 single and 5 multiple-nodule cases (1 double, 2 triple, and 2 quadroon). The size of nodules ranges from 5mm to 30mm in diameter, and are locate in the lung field only. The multi-nodule cases are represented by dark and linked locations. The single nodule cases are not linked and represented by bright localizations.

Subjects were informed of the number of radiological cases included in the image data bank and the possible number of target that they can find per case. Subjects were asked to perform a free visual search with unlimited time in order to find any nodules (targets), disregarding any other pathologies. In the process of the search, recognition and
detection tasks, each subject was instructed to mark the location of targets using a single mouse click for a particular nodule and verbally indicate their level of confidence according to a four category scale: ‘1’ probably not a lesion; ‘2’ probably a lesion; ‘3’ likely to be a lesion; and ‘4’ very likely to be a lesion. Each case analysis was terminated by the subject by pressing the space bar.

2.2 OUTCOME CLASSIFICATION AND SELECTION

Decision classification considers the relative distance between the nodule centre and the mouse click location. Correct indication of nodule location is recognized in a case when that distance is smaller than the Accepted Radius, defined as the foveal visual angle and has been found as 64 pixels. Time localization is based on a threshold value of cumulative dwell time spent on a selected ROI, of 330ms per selected site. This value seems to be long enough to recognize the duration of the covert decision. Therefore, in terms of the correctly detected nodules, visual search errors (unfixated nodules) were not included in the time order classification. Only responses with prolonged cumulative dwell time were taken as fixations. Thus, both detection and recognition events were included in the study [16]. Figure 3 shows the chart flow of the applied procedure, which considering the differences between the recognition and detection errors.

2.3 FIXATION BACKGROUND ANALYSIS

The spatial frequency analyses were performed on cases with at least one overt or FN decision. The stationary wavelet packet transforms [17] were obtained for each selected region of interest up to the 3rd level of decomposition frames by Daubechies basis functions [18], which gives 84 features in total. Each feature – called a spatial frequency band (SFB) – has physical meaning in terms of a spatial frequency and orientation unit. However, the stationary packet transform is highly redundant [19]. A statistical analysis of wavelet representation was proposed in order to find, which SFB contributes to the non-redundant description of local background properties [2]. ANOVA were used to find the significant differences (p<0.05) between the amount of energy carried by a set of spatial frequency bands, which constitutes a wavelet representation of selected ROI. In consequence, the initial wavelet representation containing 84 bands were reducing to only 29 ‘unique’ bands neglecting the redundant components in the spatial frequency domain.

3 RESULTS

The results section starts with the descriptive analysis of the number of classified outcomes following a certain type of first decisions recorded in the all cases (see 3.1). Eye tracking data characterising each decision site are presented.
according to the type of the first outcome (see 3.2). The results of the correlation analysis between the JAFROC study is presented in section 3.3. Finally, the last section (see 3.4) contains results from two different SFA.

### 3.1 NUMBER OF OUTCOMES

The number of outcomes (TP, FP and FN) gathered from each observer group in order of the previous type of decision (Figure 4).

![Figure 4](image-url)  
Figure 4. Distributions of decisions (TP, FP and FN) following from different types of first outcome (columns: TP, FP or FN as a first decision in the case) made by observers from one of the experience groups (rows: Expert, Novice2, Novice1 and Naive). Number of outcomes (count) is presented in logarithmic scale (log10).

### 3.2 EYE TRACKING FOLLOWING TP OR FP

The influence of the first overt decision has been also analyzed in order of the cumulative dwell time spent on the nodule sites classified as TP or FN and also on sites where the subject incorrectly indicated nodule presence (Figure 5). Moreover, the time of these decisions were also investigated (Figure 6).
Mean Cumulative dwell time associated fixations [ms]

Naive Novice1 Novice2 Experts

Mean Cumulative dwell time associated fixations [ms]

Outcome

First Decision in Case
- TP is 1st
- FP is 1st
- FN is 1st

Figure 5. The mean times [ms] of dwelled overt outcomes (rows) collected from 10 subjects for each group (set of bars) are presented in order of the type of the first decision made on the case TP, FP, FN (left to right order). The error bars represent standard error.

Mean Time event [ms]

Outcome

First Decision in Case
- TP is 1st
- FP is 1st
- FN is 1st

Figure 6. Mean time events [ms] of overt decisions (rows) following a TP, FP or FN (in left to right order) for each subject’s experience group (set of bars). The error bars represent standard error.
3.3 JAFROC FOM CORRELATION WITH THE NUMBER OF OUTCOEMS

The importance of the first overt decision in lung nodule detection from PA chest radiographs was found in a strong correlation (Table 1) between the JAFROC score for each subject and correctness of their first overt decision (numbers of TP or FP).

Table 1. Correlation analysis between the overall performance quality measured with JAFROC Figure of Merit (FOM) and the corresponded number of overt decisions (#TP and #FP) made on the whole image bank. The number of incorrectly indicate pulmonary presence were analyzed based on abnormal set of images (‘abnormal only’) and using data collected from all cases (‘overall’).

<table>
<thead>
<tr>
<th>experience</th>
<th>#TP</th>
<th>#FP</th>
<th>images set</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAFROC</td>
<td>0.612*</td>
<td>0.744*</td>
<td>-0.750*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>abnormal only</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.727*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>overall</td>
</tr>
</tbody>
</table>

(*) The significance of the correlation coefficients were obtained at level p<0.05.

3.4 SPATIAL FREQUENCY ANALYSIS OF ROI

The implication of the first decision was studied as a cascade effect. Spatial frequency analysis shows a significant effect of the first decision type on the perceptual criteria based on which observers might distinguish nodule regions (either classified as a TP or FN) and the overestimated nodule presence (FP). That effect manifests in a different set of SF bands used for ROI classifications depending on the results from the holistic view, whether led to TP, FP or FN outcome. For the most experience subjects, almost all bands are different for TP and FP locations when the subject made a FP as a first decision. In contrast, only 1 band can be used to distinguish these locations in situations when the first decision is TP. A similar effect between detected and missed nodules has been found in relation to the correctness of the first decision. In cases of incorrect overt outcome, both kinds of nodules seems to overlap in visual information in the SF domain, whereas, 6 SFB were found significantly different between the average wavelet representation between TP versus FN from cases where the TP was the first outcome. Some effect has been found in a pairwise comparison between

Figure 7. Mean values of confidence level (ranges from 1 to 4) obtained for overt decisions (rows) following a TP, FP or FN (L to R) and made by subjects form different experience group (set of bars). The error bars represent standard error.
over and under estimation errors. The summary of that kind of analysis conducted upon all subjects’ groups is presented in Table 2, where significant differences between wavelet representations of selected decision sites depend on the type of the first outcome.

Table 2. First decision made on a case was used as a factorization variable in the SFA of the TP, FP and FN wavelet representations calculated based on the Expert outcomes only. The number of statistical sig. (p<0.05) differences between two wavelet representation contains 29 SFB depends on the first decision at a particular case. The number indicates the scale of statistically distinguishable differences between the spatial frequency description of paired outcomes: TP x FP, TP x FN and FP x FN following a certain decision: TP, FP or FN.

<table>
<thead>
<tr>
<th>Level of experience:</th>
<th>Experts</th>
<th>Novice2</th>
<th>Novice1</th>
<th>Naives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st decision:</td>
<td>TP</td>
<td>FP</td>
<td>FN</td>
<td>TP</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>FN</td>
<td></td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td></td>
<td></td>
<td>FN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of experience:</th>
<th>Experts</th>
<th>Novice2</th>
<th>Novice1</th>
<th>Naives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st decision:</td>
<td>TP</td>
<td>FP</td>
<td>FN</td>
<td>TP</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>FN</td>
<td></td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td></td>
<td></td>
<td>FN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pairwise Comparison of SFB sets:</th>
<th>TP x FP</th>
<th>FP x FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP x FP</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>FP x FN</td>
<td>23</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison of SFB sets:</th>
<th>TP x FN</th>
<th>FP x FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP x FN</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>FP x FN</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

FN following a certain decision: TP, FP or FN.
DISCUSSION

The influence of the first decision was studied in lung nodule detection from PA chest radiographs, where subjects with different radiological experience participated in eye tracking experiments. The cascade effect [2] has been found in the radiological performance of pulmonary nodule detection from PA Chest Radiographs. The probability of the correct overt decision is greater if it is following a TP repose as a first decision made upon case. Although, this tendency is confirmed across all expertise groups, it is stronger within more experience subjects (Figure 4). A similar effect occurs for FP outcomes, which are frequent after over estimation of nodule presence in a first decision. In the case, of an unmarked nodule that is dwelled longer, it is more likely to lead to a FN decision, as a subsequent outcome.

The implication of the type of the first decision has been found in some of the eye tracking data related to outcomes following a TP or FP (Figure 5 - Figure 7). The distribution of fixation points around an image plane gives insight to the visual sampling strategy of display information. Human subjects are very selective in terms of visual attention allocation, which is measured according to the amount of time during which a particular region was foveated. This assumption that visual attention is allocated with the prolonged dwelled sites [3] has been a widely accepted paradigm in medical image perception and radiological decision-making studies [4][5]. The mean value of the cumulative dwell time spent on a certain ROI may indicate the scale of subjective difficulties. Thus, easy-to-classify cases will be characterised by a short cumulative dwell time. This aspect has been studied in terms of the implication of the first decision. Figure 5 presents the significant differences between the mean values of cumulative dwell time depending on the type of outcome and the correctness of the first hit. And so, the correct recognition of the pulmonary nodule in first place seems to reduce amount of time necessary to detect subsequent nodules. This effect can be related an uncertainty reduction on nodule containing sites. Moreover, the process can be supported that TP regions are similar in some respect and observers just tune to detect a certain type of feature. In contrast, dwell time increases on FP outcomes in the cases where TP was made in first place. The reverse effect is observed in cases with FP as a first hit. Moreover, the TP outcomes following a TP are made in general much quicker compare to those TP, which occur after FP (Figure 6). This observation supports previous arguments about the uncertainty in decision-making. It seems that observers make TP decisions in a short period of time from a previous TP, because they may find it easier to find and correctly classified a pulmonary nodule.

The final aspect related to the difficulty scale is the subject’s level of confidence in indicating a nodule presence (Figure 7). There is no significant effect of the first decision on the subject responses in all experience groups, with the one exception. The naives tend to be more confident in marking true nodule locations in cases with FP as a first hit. We have, at present, no satisfactory explanation for this observation.

Another aspect investigated in this study is dedicated to the overall performance quality in order of the influence of the first decision. The results from JAFROC analysis confirmed the first decision influence on the overall radiological performance. This result is according to our expectations. Moreover, the cascade effect seems to be even more strongly correlated with the score of the FOM than the level of experience. Thus, the correctness of the first decision implicates the whole image interpretation. That conclusion can be also supported by the early work in mammography [2] or SOS effect [11].

The implication of the first decision was studied in terms of the scale of similarities between higher order decision (second and higher) in terms of the type of first order outcome. Significant distinguishable differences in SF properties have been found between decision sites based on the effect of global impression. A similar effect was obtained compared to the recent study conducted by Mello-Thoms (2009). The importance of the Gestalt view in radiology is supported by recent studies [1], which influences the shifting paradigm in theoretical frameworks of medical image perception [2]. The global impression affects the way a radiograph is analysed and in consequence, whether subject responses are correct. A fault in the first gist is thought to determine the likelihood of recognition error made in the first overt decision. It has been shown to effect the rest of the radiological task performance in mammography [2]. The proportion of TP, FP and FN decisions following a particular type of very first response support this opinion of the effect of the global impression. The strong correlation between correct or incorrect initial recognition and the TP and FP outcome yield in the same case were found. Similarly, the proportion of FP that follow FP at each time order stage, and TP following TP is maintained by all subjects. However, more experience readers are also more resistance to FP after FP. Essentially, the proportions of the missed nodules decreased along the experience scale in cases where FP was recognized as a first decision.
Global impressions influence radiological performance via a mental bias on the target (lung nodule) locations. It establishes or indicates an existing 'preferred perception'. That influence seems to be reflected in the first overt decision [2]. Thus, the correct recognition during first hundred milliseconds turns up a TP, whereas, a fault in that process brings FP as a first decision on a case. However, it is felt that the global impression influences not only the first decision, but affects the whole radiological performance on a case. Thus, it changes or reveals the perceptual criteria held by the observer in distinguishing TP from FP, which probably affects the rate of these outcomes. That rate depends on the type of the first decision. In consequence, high quality performance occurs when TP is made as a first decision, which has been supported by the correlation with the average JAFROC FOM value.

5 CONCLUSION

The global impression which is thought to affect the first decision has been analysed in terms of the influence on the rest of performance. The type of 2nd and higher order overt decisions depends on the correctness of the very first decision in the case and also each decision corresponds to the status of the preceding one. Thus, the Gestalt view seems to influence the perceptual criteria based on which the subject distinguishes an overt decision. This interpretation shows some agreement with the conclusions presented by Mello-Thoms et al, who suggest that experts may use specific neural connections – a set of spatial frequency channels tuned to specific object detection – during visual searching in a radiological task [20][2].

The main contribution to knowledge of this work is based on the fact that for the very first time the SFA was conducted on the radiological task of lung cancer detection from PA chest radiographs. There is a degree of novelty in the methodological approach to the problem, and the investigation is entirely original in investigating the radiological modality that is most frequently performed worldwide. It is most important in terms of contribution into what is known in the psycho-physics of general perception and it provides insights into the nature of radiological expertise. The comparable results obtained in this study and presented in previous studies support the statement that even though radiological skills are not transferable from one imaging modality to other, similar principles may explain the development to expert level and the way experts perform their task. The work lends significant weight to the argument that spatial frequency channels coded through a wavelet paradigm are a characterising feature of visual perception and that this is phenomenon is generalisable to areas of radiology other than the special area of mammography. Also, this work is an extension of the previous study to the non experts group, which allows investigating the trend in radiological experience development.

6 REFERENCES


