Channelized Hotelling observer assessing microcalcification detectability on 2D mammography: a first application to study the impact of tube voltage

K T Wigati¹,², L Vancoillie¹, E Salomon¹, G Zhang⁴, L Cockmartin¹,⁴, N Marshall¹,⁴, H Bosmans¹,⁴, D S Soejoko², K Bliznakova¹, D Petrov¹

¹ Department of Imaging and Pathology, KU Leuven, 3000 Leuven, Belgium
² Department of Physics, University of Indonesia, 16424 Depok, Indonesia
³ Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, 1090 Vienna, Austria
⁴ Department of Radiology, UZ Leuven, 3000 Leuven, Belgium
⁵ Technical University of Varna, 9010 Varna, Bulgaria

Abstract. This work was aimed to assess the detectability of microcalcifications as a function of tube voltage in 2D mammography using an in-house developed Channelized Hotelling Observer (CHO). The images of a 3D structured phantom, containing acrylic beads of different diameters in water and inserted signals (microcalcifications, spiculated and non-spiculated mass models) were assessed. Images had been acquired on a Siemens Mammomat Inspiration system at 5 different tube voltage levels (28–32 kVp) under AEC setting. The detectability of microcalcifications was analyzed in terms of percentage correctly detected signals (PC) as well as diameter threshold (dtr), using the CHO and human observers. Two Laguerre-Gauss and 8 Gabor channels were included in the two-layers CHO. The model and human observer results were retrieved from a 4-alternative forced choice (4-AFC) study. At 28 kVp and 32 kVp, the diameter thresholds dtr of microcalcifications together with their 95% confidence intervals (CI95) were 0.114 [0.110-0.118] mm and 0.119 [0.116-0.122] mm for the CHO, and 0.110 [0.093-0.127] mm and 0.123 [0.111-0.134] mm for the human observers. The Pearson correlation (r) between the PC values of model and human observer was more than 0.934. The in-house developed CHO and the human observer scores correlated very well for the application on 2D digital mammography acquired with different tube voltages. The overlapping range of CI95 of the dtr for the tested kVp shows that the tube voltage setting does not significantly affect the detectability of microcalcifications neither by the CHO nor by the human observers.

1. Introduction
The main purpose of medical imaging is to assist medical experts, such as radiologists and physicians, in achieving high quality diagnostic decisions and treatment deliveries. In this regard, the quality of a medical image should be defined in terms of how well the desired information, i.e. the specific diagnostic task, can be extracted from the image by an observer. The gold standard for image quality evaluation is a Receiver Operating Characteristics (ROC), Visual Grading Analysis (VGA) or M-Alternative Forced Choice (M-AFC) study, performed by human observers [1]. Nowadays, model observers are also being considered for task based image quality assessment, in particular for quality control (QC) applications [2]. In present study, we developed a channelized Hotelling model observer (CHO) in order to investigate the impact of tube voltage on microcalcification detectability in 2D digital
mammography. This work is part of a larger project that evaluates the limiting values of a QC protocol for 2D digital mammography. For the tube voltage accuracy test, most QC protocols stipulate that the tube voltage, expressed in kVp, should be correct up to ± 1 kV and systems could be suspended from ± 2 kV onwards [3,4]. We have assessed the impact on detectability of lesions for settings at these 2 limits. Images have been assessed with both human and CHO observers. The use of an appropriate model observer instead of human observers is expected to increase performance efficiency. However, before we can apply the CHO to test the impact of tube voltage, it has to be proven that this model observer can be used for deviating tube voltage conditions. For this reason, the human observer study part was included. Surely, a well working CHO for different tube voltages would be of huge benefit when large numbers of tests of that kind have to be performed.

2. Materials and methods

2.1. The 3D structured phantom

The 3D structured phantom, as developed by Cockmartin et al. [5], was used. It is composed of a structured background and lesion-like objects. The structured background was formed of a compressed breast-shaped PMMA container of 200 mm diameter and 48 mm thickness (equivalent with breast thickness of 60 mm) that was filled with six different diameters of polymethyl methacrylate (PMMA) spheres and the rest was filled with water. Microcalcification clusters were made from particles of calcium carbonate (CaCO$_3$) and set into 5 size groups, where each size group contains 5 microcalcification specks with diameter range of 90–100, 112–125, 140–160, 180–200 and 224–250 µm. The microcalcification targets were glued on a PMMA plate (200 × 20 × 2 mm$^3$) and placed in the middle of the phantom at 50 mm from the chest wall side (Figure 1).

2.2. Image acquisition

The 2D images of the 3D structured phantom were acquired on Siemens Mammomat Inspiration mammography system. Five different kVp values were selected. The kVp setting was initially determined for 48 mm of thickness without any additional spacers and resulted in a tube voltage of 30 kVp with W/Rh anode/filter. The other four kVp values were set manually: 28, 29, 31, 32 kVp at the same anode/filter. Ten images were acquired for each kVp setting and the phantom was shaken after each acquisition in order to configure multiple variants of the image background. Under semi-automatic exposure control mode, the mAs values were 129 ± 3; 115 ± 4; 98 ± 2; 83 ± 3 and 74 ± 2, for 28 to 32 kVp respectively (0.9–1.1 mGy).

2.3. Image data set

“For presentation” type images were used in this study. The phantom images had dimensions of 239.36 × 304.64 mm$^2$, with resolution of 0.085 mm/pixels. Twenty regions of interest (ROIs) with dimensions of 20 × 20 mm$^2$ (236 × 236 pixels) were segmented from a single phantom image: 15 ROIs were segmented from signal free areas within the phantom and 5 ROIs with the calcification clusters located in the center. Samples of signal-absent images and signal-present images for each kVp setting are shown in Figure 2.

Figure 1. (a) Physical 3D structured phantom, (b) Part of phantom region containing microcalcification targets (the white dots represent the glue) [5], and (c) 2D mammography image of the 3D structured phantom.
2.4. Human observer study
The observer study was conducted in a 4-AFC paradigm. In such reading experiment, four images are presented simultaneously to the observer, where three of them are coming from a pool of signal absent images and one from signal present images. The observer’s task is to select the signal-present image out of the four. This 4-AFC trial is repeated for all available signal present images in order to obtain the percentage correctly detected targets (PC), determined by the number of correct decisions over the number of 4-AFC trials (1).

\[ PC = \frac{\text{number of correct decisions}}{\text{number of 4-AFC trials}} \]  

(1)

In present study, for each target diameter ten trials were performed. The PC value as a function of diameter was then fitted with a psychometric curve to obtain the diameter threshold. This psychometric curve was based on a logistic function, which starts at 0.25 (guess level) and increases to 1 (when all targets are correctly detected) in a sigmoidal manner (2):

\[ PC(d) = 0.25 + \frac{0.75}{(1+e^{-f(d-d_{tr})})} \]  

(2)

in which \( d \) is diameter, \( d_{tr} \) is diameter threshold, and \( f \) is a free parameter to be fitted with the psychometric curve. The parameter \( f \) affects how steeply the function rises as it passes through its midpoint (PC = 0.625), where the \( d_{tr} \) is determined. [6,7].

Five medical physicists participated in this observer study. The 4-AFC analysis was performed on a 5 MP monitor calibrated to the DICOM GSDF standard (Barco Coronis), utilized with in-house software for image assessment and statistical evaluation [8]. The observers read the images at approximately 40 cm away from the monitor and were allowed to adjust contrast and brightness to reach the optimum condition for observation.

2.5. Channelized Hotelling model observer
The CHO algorithm used in this study was adapted and developed from earlier studies on digital breast tomosynthesis (DBT), described in detail in Petrov et al. [9], but a summary will be presented here.

The test statistic of a channelized Hotelling observer can be calculated using the following formula [10]:

\[ t_{CHO} = w_{CHO}^T v \]  

(3)

\[ v = U^T g \]  

(4)

\[ w_{CHO} = s^T K_p^{-1} \]  

(5)

in which \( v \) is the channel output vector, which is a product of the channel matrix \( U \) and the image vector \( g \) (4). The observer template \( w_{CHO} \) is formed from the expected signal \( s \) and the inverse of the covariance matrix \( K_p^{-1} \) (5). Due to insufficient training images, the expected signal was approximated.
by a Gaussian function with full width at half maximum (FWHM) equal to the average diameter of the calcification specks, and the covariance matrix was estimated from 150 signal absent images.

Given the CHO template properties, the CHO requires the exact locations of the calcification particles forming the cluster. Due to manufacturing tolerances in the production of the phantom and always different positions of the phantom on the bucky, the locations are not known with high accuracy. Therefore, prior to estimate the detectability of the cluster, a localization algorithm had to be included, i.e. the CHO was split into localization and quantification stages. The localization stage employed a CHO with two Laguerre-Gauss (LG) channels by Gallas and Barrett [10], shifted in specific locations within the ROI producing a test statistic maps containing the probability for a present calcification particle. Then the locations with the five maximum test statistics were stored and reused in the subsequent quantification stage. The quantification stage used a CHO with eight Gabor channels by Petrov et al. [11]. The rationale behind this approach was to take advantage of the tuning ability of the Gabor channels, without compromising the localization of the particles. This way the calcification speck locations were estimated with high accuracy using the ideal LG channels, and the cluster was detected using the anthropomorphic Gabor channels [12]. The tuning of the channel parameters used the same basic principle: the LG channels were tuned to maximize the localization properties of the algorithm and the Gabor channels were tuned to maximize the CHO agreement with the human observer scores.

In order to compare both human and model observer results, the CHO scores were expressed as PC in a 4-AFC study. In order to do that, 3 signal-absent test statistic values were compared to one signal-present test statistic value. A right decision is deemed if the latter has the highest value of all four.

3. Results

3.1. Percentage correctly detected signals and model observer validation
Percentage correctly detected signals by human observers and the CHO model for each target diameter separately are shown in Figure 3. As expected, PC decreased with decreasing target diameter. For the 2 largest types of microcalcifications (180–200 µm and 224–250 µm), the human observers and the CHO model gave identical absolute PC values (1.00 ± 0.00) for all kVp settings. For the other microcalcification groups, most of the PC values by the CHO model were found to be lower than the ones by the human observers.

![Figure 3. Percentage correctly detected signals as a function of tube voltage for the different microcalcification groups](image-url)
Despite rising uncertainties on PC values, the group of microcalcifications of 140–160 µm was clearly visible by human observers and the CHO model observer, indicated by the range of PC values between 0.80 ± 0.03 and 1.00 ± 0.00. The PC values for microcalcifications between 112–125 µm were in the range of 0.56 ± 0.09 and 0.84 ± 0.05 by the human observers, and between 0.47 ± 0.05 and 0.67 ± 0.03 by the CHO model. The average PC values for the smallest microcalcifications (90–100 µm) were found around the guess level of the 4-AFC analysis (PC = 0.25), i.e. 0.27 ± 0.16 for the human observers and 0.23 ± 0.04 for the CHO model.

The Wilcoxon signed-rank test between the human observers and the CHO model resulted in p = 0.3750, p = 0.1250 and p = 0.1250 for observation of microcalcifications in the ranges 90–100 µm, 112–125 µm, and 140–160 µm respectively. Scatter plots of the PC values for all of target diameters and kVp tested are shown in Figure 4. The minimum Pearson correlation (r) was 0.934 at 29 kVp. These values indicated that the CHO decisions were in high agreement with the human observer.

3.2. The impact of tube voltage on microcalcification detectability

Figure 5 (a) shows the psychometric curve of the PC values as a function of microcalcification diameter. Figure 5 (b) shows the diameter thresholds resulting from Figure 5 (a) at each kVp setting. The diameter thresholds together with their 95% confidence interval (CI95) from 28 to 32 kVp were respectively 0.110 [0.093-0.127] mm, 0.117 [0.109-0.126] mm, 0.118 [0.107-0.129] mm, 0.118 [0.105-0.131], and 0.123 [0.111-0.134] mm for the human observer; and 0.114 [0.110-0.118] mm, 0.124 [0.118-0.130] mm, 0.120 [0.115-0.124] mm, 0.135 [0.129-0.140] mm, and 0.119 [0.116-0.122] mm for the CHO model. The overlapping ranges of the confidence intervals for the tested kVps show that the tube voltage setting does not significantly affect the detectability of microcalcifications, neither by the CHO model nor by the human observer.

4. Discussion and conclusion

This work examined detectability of microcalcifications in 2D mammography by an in-house developed CHO and by human observers for different tube voltage settings. The range of tested tube voltages varied from 28 kVp until 32 kVp and centered on the AEC obtained value of 30 kVp. For this range, most QC
protocols would require an accuracy that is better than 1 kVp, where occasionally a deviation of 2 kVp would even be used to suspend a system. Within the range of tested kVps, the PC values of the five different microcalcification diameters with their CI95 were found overlapping, either observed by CHO or human. We can conclude that deviations in tube voltage of ±1 and ±2 kV did not affect image quality figures for detection of calcifications and masses.

We successfully developed a model observer for 2D digital mammography applications. The selection of channels and tuning parameters (2 LG channels and 8 Gabor channels) was based on their performance in terms of signal localization and quantification. The final choice was the set that provided the best Pearson correlation between CHO and human readings $(r > 0.934)$. It must be noted, that the CHO was trained and applied on the same image dataset, due to an insufficient amount of images. Whether the same CHO could now be used on other systems or for an even wider kVp range or on other systems remains to be proven.

The imperfect correlation between human and CHO observers may be due to several factors. (1) Difference in scoring method between the CHO and the human observer. (2) Limited accuracy of image positioning: in order to localize the signal, one reference point was selected (the center microcalcification of the biggest group). The use of two reference points may give a better accuracy than one reference point. (3) Limited number of images: increasing the image sample could reduce the uncertainties as well as help to achieve statistically robust data. In this study we used a limited amount of images, and images used for training were subsequently re-used to obtain the scores. (4) Limited number of tested channel selections and tuning factors.

This study was continuation of CHO work mainly performed in breast tomosynthesis but also in 2D FFDM and correlation with human readers was deemed acceptable, making the CHO a promising candidate for future applications.

References
[1] Dance D R, Christofides S, Maidment A D A, McLean I D and Ng K H 2014 Diagnostic radiology physics: a handbook for teachers and students (IAEA)
[2] Ba A et al. 2018 Med. Phys. 45 (7) pp 3019–3030
[3] European Commission 2006 European guidelines for quality assurance in breast cancer screening and diagnosis 4th ed.
[4] European Commissin 2012 Radiation Protection No 162
[5] Cockmartin L et al. 2017 Phys. Med. Biol. 62 pp 758–780
[6] Young K C et al. 2008 Evaluation of software for reading images of the CDMAM test object to assess digital mammography system Proc. SPIE 6913
[7] Klein S A 2001 Perception & Psychophysics 63 pp 1421-1455
[8] Zhang G, Cockmartin L, and Bosmans H 2016 A four-alternative forced choice (4AFC) software for observerperformance SPIE
[9] Petrov D et al. 2019 Model and human observer reproducibility for detection of microcalcification clusters in digital breast tomosynthesis images of a 3D structured test object JMI
[10] Gallas B D and Barrett H 2003 J Opt Soc Am A Opt Image Sci Vis 20 (9) pp 1725–1738
[11] Petrov D et al. 2018 Model and human observer reproducibility for detecting microcalcifications in digitalbreast tomosynthesis images Proc. SPIE 10577
[12] Barret H H, Yao J, Rolland J P and Myers K J 1993 Proc. Natl. Acad. Sci. USA Vol 90 pp 9758 9765

Acknowledgement
This work is part of project on “CHO Model Observer Workshop” held in Leuven 2018, funded by MaXIMA project (a European Union’s Horizon 2020 research and innovation program), and P3-Stroke (a European Union Horizon2020 EIT Health project). The authors would also like to acknowledge Janne Vignero and Michiel Dehairs for participating in the human observer study.