

Effects of Medical Display Luminance, Contrast and Temporal Compensation on CHO Detection Performance at Various Browsing Speeds and on Digital Breast Tomosynthesis Images

Cédric Marchessoux¹, Ali N. Avanaki^{1,*}, Predrag R. Bakic²,
Tom R.L. Kimpe¹, and Andrew D.A. Maidment²

¹ BARCO N.V., Healthcare Division, Kortrijk, Belgium
ali.avanaki@barco.com

² University of Pennsylvania, Department of Radiology, Philadelphia, PA, USA

Abstract. Prior studies have shown that temporal compensation of medical displays improve the performance in detecting lesions for digital breast tomosynthesis (DBT). This has been proven both by using computer simulations as well as clinical experiments. This paper, by using computer simulations, studies (i) the effect of the maximum luminance (L_{\max}) and contrast (L_{\max}/L_{\min}) of the medical display on lesion detection performance, and (ii) the effect of temporal compensation of the display (by comparing displays with and without this feature) on lesion detection performance, with several slice browsing speeds using a fractional frame repeat (FFR) scheme to model displays' behavior when the refresh rate is not an integer multiple of the browsing speed.

Keywords: Medical display, digital breast tomosynthesis (DBT), clinical studies, channelized Hotelling observers (CHO), browsing speed.

1 Introduction

Digital Breast Tomosynthesis (DBT) is a three-dimensional imaging technology that involves acquiring images of a stationary compressed breast at multiple angles during a short scan. The individual images are then reconstructed into a series of thin high-resolution slices that can be displayed individually or in a dynamic mode. Because reviewing images for this modality typically is done in a dynamic mode, which was not the case with full field digital mammography, and because breast cancer screening requires the best image display quality, a display optimized for DBT modality was developed (BARCO MDMG 5221). This display was optimized with improved intrinsic key characteristics such as contrast, luminance and temporal response. This article presents the result of a follow up study reported in [1]. Additional browsing speeds and display parameters such as contrast and luminance are considered and reported. In addition, input generated by a voxelized breast model [2] is also used. This is part of

* Corresponding author.

virtual clinical trial (VCT) platform that is under development, with the objective to compare real DBT images in which artificial lesions were introduced and fully simulated DBT images. For fully managing a VCT, every step in the chain should be controllable. Hence, using a phantom generator in VCT simulations to provide a customized input is necessary.

1.1 Prior Work

In [1], DBT reconstructed slices were used in which single micro calcifications were inserted. We have available to us a compiled version of a commercially used DBT reconstruction engine along with anonymized and pre-processed projection data (P) from real patients and geometry information necessary to reconstruct those projections using the DBT reconstruction engine. By pre-processed projections, we mean that any vendor/device specific projection processing such as bad pixel correction and beam hardening correction have already been performed on the raw projection data and that projection data can be now used as input to a known reconstruction algorithm. The DBT reconstructed images we reviewed consist of about 50 slices, each on a 1200 x 2400 matrix. The DBT voxel size was approximately 0.1 mm x 0.1 mm x 1.0 mm. A display with improved features such as temporal response compensation was clinically evaluated with the use of a numerical observer described in [3]. The numerical observer is an extension of a Channelized Hotelling Observer (CHO) for multiple slices that can be applied for quantifying the effect of the browsing speed of a system on lesion detection performance. A multi-reader multi-case (MRMC) analysis [4] was performed with 5 readers, each trained with 500 image pairs, and all reading the same 500 test image pairs. Only integer frame repeats (FR) were used that correspond to slice browsing speeds of F_{refresh}/FR (50, 25, 50/3, 12.5, 10 slice per second, for F_{refresh} of 50 frames per second) which is very limiting as on real displays that browsing speed is desired to be changed continuously. A sample slice and a 1-D plot (central pixel luminance over slice number) are shown in Figure 1.

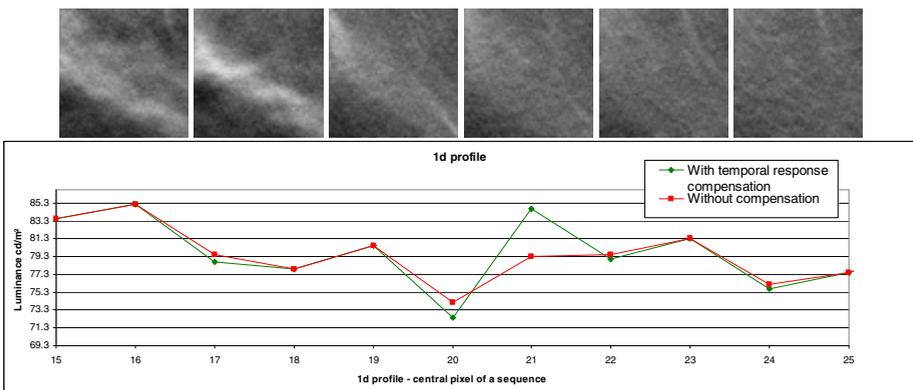


Fig. 1. Example of cropped (64 x 64 pixels) DBT slices from a 41-slice stack with a 0.4 mm lesion inserted in the center and a 1-D plot central pixels through 11 slices in cd/m^2 for temporal-response compensated and uncompensated displays. The signal is inserted in slice 21.

2 Methods

A key prerequisite of excellent system design in imaging systems is the control of the interplay of all its elements. Typical elements are the technology of image capture, the representation of the images as digital data, processing or enhancing of these data for a specific image display, the nature of the display technology (print or softcopy), and the psychometric judgment of the images through a human visual observer model [5]. An integrated approach, which combines a complete system model for a given imaging technology and a human visual observer model in one computational workbench, does therefore represent a great improvement for systematic image system design, optimization and even simulation of technology feasibility prior to prototypes. Engeldrum proposed a methodology named the Image Quality Circle [6], which shows the different phases to control and simulate a complete chain in the domain of vision. It also shows the links and relations between technology variables that we control from a product to the physical image parameters that we get from system modeling. The resulting image quality should be correlated to and optimized for the human perception or customer perceived preferences. From technology variables to user's preferences, the circle covers the complete chain. This methodology was used to develop and optimize a medical display for the digital breast tomosynthesis modality. A C++ simulation platform called MEVIC (Medical Virtual Imaging Chain) was used for simulating the complete chain from the image capture until the visualization of the images [7]. The virtual medical imaging chain starts with simulation of the image acquisition, over a hardware and software image processing pipeline and ends with the visualization by the medical specialist on the image display. The aforementioned chain is modeled as a cascade of three main modules: the virtual image capture, the virtual display and the virtual observer.

The key techniques that are used in MEVIC simulations for the current study are briefly described in the remainder of this section.

2.1 Simulation of DBT Images

As described in the prior work, reconstructed DBT slices from a real acquisition device were used as input images to the virtual imaging chain. In total, 6000 cropped 64x64-pixel 41-slice stacks were used. The pixel values are coded in 10 bits. This dataset have two categories: healthy and diseased. The synthesized 3D mass breast or micro calcification lesion of a given density is inserted in the reconstructed background volume. In comparison with the original images in [1], the input images are modified to have the maximum possible contrast (covering the full span of [0, 1023]) with a single offset/gain transform in each stack. This corresponds with clinical practice to use contrast-enhancement and window-level settings to maximally make use of the available grayscales of the medical display.

As the second objective to this study, we want to use the artificial backgrounds generated by anthropomorphic software breast phantom developed at the University of Pennsylvania [8]. It simulates the breast anatomy based upon the detailed analysis of histological and radiological images. The arrangement of breast tissues at the large

and medium spatial scales is realistically simulated using a region growing approach. Synthetic x-ray images of the phantom are generated by simulating the breast deformation during the mammographic compression using a finite element model proposed in [9], followed by a model of the x-ray projections of the compressed phantom, assuming mono-energetic x-rays without scatter. 1648 stacks, each consisting of 32 64x64-pixel slices of phantom were used. The lesions are inserted using the procedure described in [1] to make the diseased stacks. The stacks are then randomly re-ordered to make it less likely that corresponding healthy and diseased images fall into the same training or test sets.

2.2 Display Simulation Chain

2.2.1 Contrast and Luminance

The native curve of the display is used for factoring in the effect of contrast and luminance of the display in MEVIC. Native curve value for a certain digital drive level (DDL) is the measured luminance of the display when a certain DDL is applied at the input for a long time. L_{\max} , the maximum luminance of the display, is reached when the largest DDL (e.g. 1023, for a 10-bit display) is applied. L_{\max} / L_{\min} is the contrast with L_{\min} being the minimum luminance of the display, reached for DDL of 0 ($L_{\min}=1.05\text{cd/m}^2$ and $L_{\max}=1000\text{cd/m}^2$). The luminance values correspond to a BARCO MDMG 5221 medical display.

2.2.2 Temporal Compensation of Display

The temporal response improvement is a proprietary solution from Barco (US Patent Application No: 2010/0207,960, 'devices and methods for reducing artifacts in display devices by the use of overdrive'). This solution allows the display to reach gray intensity values within one frame time without enhancing temporal noise or introducing artifacts. This technology was integrated in a FDA approved display optimized for digital breast tomosynthesis.

2.2.3 Fractional Frame Repeat (FFR)

The ability to continuously adjust the browsing speed is a desirable feature. Using integer frame repeats, simulations will be limited as described below. Let F_{browse} show the slice browsing speed, F_{refresh} show the frame refresh rate (a display property in Hz), and FR show frame repeat. $F_{\text{browse}} = F_{\text{refresh}}/\text{FR}$. For example, at F_{refresh} of 50 frame per second (fps), if each slice is fed twice to the display at consecutive refreshes (FR = 2), the apparent slice browsing speed is $50/2 = 25$ slice per second (sps). In other words, $\text{FR} = F_{\text{refresh}}/F_{\text{browse}} = 50/25 = 2$. Hence, the browsing speeds that can be simulated with integer FRs are very limited.

By allowing a fractional frame repeat, one can have arbitrary browsing speeds as follows. As an example, F_{browse} of 40 sps can be achieved if we make 5 frames out of every 4 slices. In this case $\text{FR} = F_{\text{refresh}}/F_{\text{browse}} = 50/40 = 5/4$. To that end, we use an error accumulation method to find out which slices should be repeated: starting from the beginning of the stack (the residue is initially set to zero), each slice is copied floor(FR+residue) times, generating that many frames, and the residue is updated to

FR+residue-floor(FR+residue). This way, when the residue goes above one, an extra frame with a copy of the current slice is inserted. To have a slice browsing speed of 40 sps, on a 41-slice stack (comprised of slices 1, 2, ..., 41), when $F_{refresh}$ is 50 fps, the following slices are written to the frame buffer: 1 2 3 4 4 5 6 7 8 8 9 10 11 12 12 13 14 15 16 16 17 18 19 20 20 21 22 23 24 24 25 26 27 28 28 29 30 31 32 32 33 34 35 36 36 37 38 39 40 40 41. In this example, slice n is copied twice if $\text{mod}(n, 4) = 0$, and all other slices are copied only once.

2.3 Multi-slice Channelized Hotelling Observer (msCHO)

The multi-slice Channelized Hotelling model Observer (msCHO) described in [1, 3] is used with 10 LG channels of spread 15 for both real and artificial background data. The msCHO performance is computed for the pixel values achieved at the end of each refresh cycle during the T_{browse} . For example, when the frame repeat $FR = 3$ (see Tables 1 and 2), the detection performance is computed for image content at the end of each $1 \times T_{refresh}$, $2 \times T_{refresh}$ and $3 \times T_{refresh}$. Our observer only uses the three central slices of each stack as the other slices are lesion-free.

3 Results

3.1 Results on Real DBT Reconstructed Slices

The results on real DBT background slices are reported in Table 1. They show that FFR is working as expected since the AUCs for the FFR-generated browsing speed are similar to AUCs for speeds generated by regular (integer) frame repeating. Table 2, reports the same for a display without temporal compensation.

Table 1. Detection performance on real DBT reconstructed slices for 2 FFRs (30 & 40 sps) and three integer frame repeats (16.67, 25 & 50 sps) for a temporally compensated display on contrast-stretched data. The computations are performed in an MRMC study with $N_{rd} = 5$ readers, each trained with an independent subset of $N_{tr} = 500$ image pairs and all reading the same test set of $N_{ts} = 500$ test image pairs. The size of the ROI is 3. The AUCs and standard deviations are calculated using the one-shot method [4].

	Which after LCD frame is used to train 2D-CHO?	FR=1	FR=50/40	FR=50/30	FR=2	FR=3
AUC ± std	Frame 1	0.800 ±0.014	0.801 ±0.014	0.800 ±0.014	0.800 ±0.014	0.800 ±0.014
	Frame 2	N/A	N/A	N/A	0.801 ±0.014	0.801 ±0.014
	Frame 3	N/A	N/A	N/A	N/A	0.801 ±0.014
	Average	0.800 ±0.014	0.800 ±0.014	0.800 ±0.014	0.801 ±0.014	0.801 ±0.014

To study the effect of luminance and contrast on detection performance, we simulated two displays: (i) a low-contrast (LC) display with the same L_{\max} as that of MDMG 5221 but a 50% lower contrast, and (ii) a low-luminance (LL) display with the same L_{\max}/L_{\min} as that of MDMG 5221 but a 50% lower L_{\max} . The simulated detection performance of these displays at FR=1 are both less than 1% different than MDMG 5221.

Table 2. Detection performance on real DBT reconstructed slices for 2 FFRs (30 & 40 sps) and three integer frame repeats (16.67, 25 & 50 sps) for a display without temporal compensation. The settings are given in Table 1 caption.

	Which after LCD frame is used to train 2D-CHO?	FR=1	FR=50/40	FR=50/30	FR=2	FR=3
AUC ± std	Frame 1	0.607 ±0.026	0.607 ±0.026	0.656 ±0.022	0.608 ±0.026	0.607 ±0.026
	Frame 2	N/A	N/A	N/A	0.800± 0.014	0.801 ± 0.014
	Frame 3	N/A	N/A	N/A	N/A	0.801 ± 0.014
	Average	0.607 ±0.026	0.607 ±0.026	0.656 ±0.022	0.704 ±0.020	0.736 ±0.018

3.2 Results on Artificial DBT Reconstructed Slices

In Table 3, results of our preliminary experiments with the dataset generated from a sample simulated breast phantom [8] are presented: the detection performance with its standard deviation is listed for a temporally compensated display at four browsing speeds.

When a small (3%) subset of stacks with non-stationary background is added to the dataset, the AUCs drop by about 2%.

Table 3. Detection performance on artificial DBT reconstructed slices for 2 FFRs (30 & 35 sps) and two integer frame repeats (25 & 50 sps) for a temporally compensated display. The computations are performed in an MRMC study with $N_{rd} = 3$ readers [4], each trained with an independent subset of $N_{tr} = 412$ image pairs and all reading the same test set of $N_{ts} = 412$ test image pairs. The size of the ROI is 3. The AUCs and standard deviations are calculated using the one-shot method.

	Which afterLCD frame is used to train 2D-CHO?	FR=1	FR=50/35	FR=50/30	FR=2
AUC ± std	Frame 1	0.833 ±0.014	0.827 ±0.015	0.854 ±0.013	0.835 ±0.023
	Frame 2	N/A	N/A	N/A	0.853 ±0.013

4 Discussion

The insertion of single micro calcifications lesions is not completely accurate. Mimicking the x-ray absorption and generating anisotropic 3D shapes is, however, more realistic than simply inserting a simple 3D Gaussian sphere as a signal such as it is done in numerous model observer studies. The former was successfully used several times in past studies [1]. In future, within the VCT framework, lesions will be generated during the creation of the phantom.

The display MDMG 5221 used in this study features temporal response compensation. The target luminance values are reached within one frame whereas for a display without this feature, such as MDMG 5121, two to three frames are needed for final luminance values to be reached. This can be also observed in Table 1: the detection performance remains the same (within the double standard deviation range) no matter which refresh is fed to the observer. On the other hand, for an uncompensated display (Table 2), when the first frame of the slice is fed to the observer the detection performance is significantly lower. Also, as observed in Table 1 and Table 2, with the current model observer, one cannot achieve a higher AUC just by increasing the frame repeat. That is because the frames fed to the observer become almost the same after the second refresh in the display with or without temporal compensation.

In this paper and in [1], a multi slice CHO is used by computing scores for image content at the end of each $1 \times T_{\text{refresh}}$, $2 \times T_{\text{refresh}}$ and $3 \times T_{\text{refresh}}$. An alternative to this is feeding all ROI frames (those that may have part of lesion in them) to the model observer. This approach will generate results that are less consistent with those reported in [1]. Nevertheless, the average of AUCs from different refresh values mimics, in a sense, the visualization of the different frames by the observer. Further investigation will take into account the continuous light transition instead of discrete luminance values that are currently used, as well as properties of the human visual system (e.g., temporal contrast sensitivity function) in the observer model.

Detection must be performed in JND domain rather than luminance. Typically, the AUCs calculated in luminance domain are slightly lower (about 1%) than the results reported in Section 3-table 3.

Larger fluctuations in the AUC values for the experiment with the artificial dataset (Section 3.2, Table 3), as compared to the corresponding results for real data (Table 1), may be attributed to the facts that (i) fewer images are used in the experiment and/or (ii) the source of all images used in the experiment is the same phantom; thus the images have less variety. Also note that there is no significance in the fact that AUC values in Table 3 are generally larger than those in Table 1 and Table 2. This difference is a undesired side effect of the lesion insertion process: the insertion density is changed until the AUC becomes around 80%, making the classification of the set an average task (not too difficult or too easy).

We observed that the effect of temporal compensation is considerably higher than those of increasing luminance or contrast. This observation is against clinical studies with human observers and is another indication that the model observer must be improved to be a better representative of human observers. Such improvements may be achieved, as mentioned earlier, by integrating the properties of human visual system with the model observer.

5 Conclusion

In this study we gained a better understanding of current capabilities and limitations of channelized Hotelling observers to be used in virtual clinical trial framework.

Acknowledgments. This work is supported by the US National Institutes of Health (R01 grant #CA154444). Special thanks goes to Ljiljana Platiša from the University of Ghent for her support.

References

1. Platisa, L., Marchessoux, C., Goossens, B., Philips, W.: Performance evaluation of medical LCD displays using 3D channelized Hotelling observers. In: Proc. SPIE Medical Imaging (2011)
2. Bakic, P.R., Zhang, C., Maidment, A.D.A.: Development and Characterization of an Anthropomorphic Breast Software Phantom Based upon Region-Growing Algorithm. *Medical Physics* 38, 3165–3176 (2011)
3. Platisa, L., Goossens, B., Vansteenkiste, E., Badano, A., Philips, W.: Using channelized hotelling observers to quantify temporal effect of medical liquid crystal displays on detection performance. In: Proc. SPIE Medical Imaging (2010)
4. Gallas, B.: One-shot estimate of MRMC Variance: AUC. *Academic Radiology* 13, 353–362 (2006)
5. Marchessoux, C., Jung, J.: A virtual image chain for perceived image quality of medical display. In: Proc. SPIE Medical Imaging (2006)
6. Engeldrum, P.: Image quality modeling: Where are we? In: Proc. of PICS 1999: Image Processing, Image Quality, Image Capture, Systems Conference, pp. 251–255 (1999)
7. Marchessoux, C., Kimpe, T., Bert, T.: A virtual image chain for perceived and clinical image quality of medical display. *J. of Display Technology* 4, 356–368 (2008)
8. Zhang, C., Bakic, P.R., Maidment, A.D.A.: Development of an anthropomorphic breast software phantom based on region growing algorithm. In: Proc. SPIE Medical Imaging (2008)
9. Ruiter, N.V., Zhang, C., Bakic, P.R., Carton, A.-K., Kuo, J., Maidment, A.D.A.: Simulation of tomosynthesis images based on an anthropomorphic software breast tissue phantom. In: Proc. SPIE Medical Imaging (2008)