**Design and Validation of Realistic Breast Models for
Use in Multiple Alternative Forced Choice Virtual
Clinical Trials**

Premkumar Elangovan¹, Alistair Mackenzie², David R. Dance^{2,3}, Kenneth C. Young^{2,3}, Victoria Cooke⁴, Louise Wilkinson⁵, Rosalind M. Given-Wilson⁵, Matthew G. Wallis⁶ and Kevin Wells¹

¹Medical Imaging Group, Centre for Vision, Speech, and Signal Processing, University of Surrey,
Guildford, GU2 7XH, United Kingdom

²National Coordination Centre for the Physics of Mammography (NCCPM), Royal Surrey County
Hospital, Guildford GU2 7XX, United Kingdom

³Department of Physics, University of Surrey, Guildford, GU2 7XH, United Kingdom

⁴Jarvis Breast Screening and Diagnostic Centre, Guildford GU1 1LJ, United Kingdom

⁵Department of Radiology, St. George's Healthcare NHS Trust, Tooting, London SW17 0QT, United
Kingdom

⁶Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2
0QQ, United Kingdom and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ,
United Kingdom

E-mail: p.elangovan@surrey.ac.uk

Short title: Synthetic breast model and validation

Abstract

A novel method has been developed for generating quasi-realistic voxel phantoms which simulate the compressed breast in mammography and digital breast tomosynthesis (DBT). The models are suitable for use in virtual clinical trials requiring realistic anatomy which use the multiple alternative forced choice (AFC) paradigm and patches from the complete breast image. The breast models are produced by extracting features of breast tissue components from digital breast tomosynthesis (DBT) clinical images including skin, adipose and fibro-glandular tissue, blood vessels and Cooper's ligaments. A range of different breast models can then be generated by combining these components. Visual realism was validated using a receiver operating characteristic (ROC) study of patches from simulated images calculated using the breast models and from real patient images. Quantitative analysis was undertaken using fractal dimension and power spectrum analysis. The average areas under the ROC curves for 2D and DBT images were $.51 \pm .06$ and $.54 \pm .09$ demonstrating that simulated and real images were statistically indistinguishable by expert breast readers (7 observers); errors represented as one standard error of the mean. The average fractal dimensions (2D, DBT) for real and simulated images were $(2.72 \pm .01, 2.75 \pm .01)$ and $(2.77 \pm .03, 2.82 \pm .04)$ respectively; errors represented as one standard error of the mean. Excellent agreement was found between power spectrum curves of real and simulated images, with average β values (2D, DBT) of $(3.10 \pm .17, 3.21 \pm .11)$ and $(3.01 \pm .32, 3.19 \pm .07)$ respectively; errors represented as one standard error of the mean. These results demonstrate that radiological images of these breast models realistically represent the complexity of real breast structures and can be used to simulate patches from mammograms and DBT images that are indistinguishable from patches from the corresponding real breast images. The method can generate about 500 radiological patches ($\sim 30\text{mm} \times 30\text{mm}$) per day for AFC experiments on a single workstation. This is the first study to quantitatively validate the realism of simulated radiological breast images using direct blinded comparison with real data via the ROC paradigm with expert breast readers.

Keywords

Virtual clinical trials, breast model, breast phantom, synthetic breast, image simulation, 2D-mammography, digital breast tomosynthesis.

1. Introduction

2D-mammography is the predominant method for screening breast cancer owing to good diagnostic performance, low cost, short exposure time and low radiation dose (Karim-Kos *et al* 2008, Hevie *et al* 2014). However, rapid advances in technology have resulted in the introduction of digital breast tomosynthesis (DBT) into the screening process (Diekmann and Bick 2007) which has shown potential for improved cancer detection. In DBT, a series of X-ray projections are acquired at limited angles and reconstructed into image planes parallel to the detector (Niklason *et al* 1997). Some studies have shown DBT to be superior in detecting certain types of cancers (Gilbert *et al* 2016), and hence DBT has been deployed in some centres alongside 2D digital mammography for cancer detection and diagnosis.

Clinical trials with human subjects are the established method of assessing the impact of any new technology/technique in terms of diagnostic benefit. However, conventional clinical trials of diagnostic imaging techniques using clinical images can take several years to conclude, by the end of which, it is likely that a potentially improved version of the original technology/technique may be available. Large scale trials also demand significant investment in clinical resources, participant recruitment and compliance. In the context of breast cancer screening using radiological techniques, where participants are largely healthy individuals, there are also ethical issues of increased absorbed dose to trial participants.

To address these issues, virtual clinical trials (VCT) are an emerging complementary approach to conventional clinical trials for breast cancer screening (eg. Bakic *et al* 2002a, 2002b, Maidment 2014, Elangovan *et al* 2014, 2016). In VCTs, computerized modelling tools are used to simulate radiological images comparable to their real clinical counterparts. These are produced by simulating cancer pathology (Shaheen *et al* 2010, Rashidnasab *et al* 2013a, 2013b) which may be either inserted into clinical images (Elangovan *et al* 2014) or inserted into a complete simulated breast model. Various radiological image formation and degradation processes (Mackenzie *et al* 2012, 2014) are then used to model the imaging technology or technique under consideration.

To ensure the results of any VCT are comparable to those of the equivalent conventional clinical trial, all components of the VCT process need to be validated for clinical realism. Once such tools are available, VCTs provide a means for rapidly undertaking a variety of quantitative assessment tasks such as image quality, target detection and observer performance (Gong *et al* 2006, Young *et al* 2013, Kiarashi *et al* 2015).

However, to date, there has not been a method available which can generate virtual breast anatomy that is validated for visual realism by expert radiological observers. This presents a potential impediment to the widespread deployment of the VCT paradigm in breast imaging. Previous methods to simulate such breast anatomy have involved using mathematical simulation methods (Bliznakova *et al* 2003, 2010, 2012, Bakic *et al* 2011, 2014, Graff 2016, Pokrajac *et al* 2012), patient CT data (Li *et al* 2009, Kiarashi *et al* 2016, Ikejimba *et al* 2016), mastectomy CT data (Hoeschen *et al* 2005, O'Connor *et al* 2008) and computer generated noise models (Rolland *et al* 1997, Bochud *et al* 1999, Heine *et al* 1999, Castella *et al* 2008). However, it is a very challenging task to simulate realistic radiological images calculated using such mathematically derived models, since a proper selection of geometrical parameters of the basic primitives (Bliznakova *et al* 2003) and reduction of geometric appearance of the border between different tissues (Bakic *et al* 2011) are required to achieve high levels of clinical realism. Those based on breast CT images are anatomically correct for large scale features but the images lack sufficient resolution to accurately depict the fine details associated with the linear structures such as the ductal network, blood vessels and Cooper's ligaments.

As part of the development of a validated toolbox for VCTs in mammography (the OPTIMAM¹ toolbox), we have previously successfully simulated synthetic pathological structures such as mass lesions (Rashidnasab *et al* 2013a, 2013b) and micro-calcifications (Shaheen *et al* 2010), and inserted these into real clinical images using a validated image simulation framework (Elangovan *et al* 2014) to produce hybrid radiological images for VCTs. In this paper, we describe the latest addition to the toolbox, a new fast method for creating realistic 3D models of the compressed breast into which synthetic pathology can be inserted prior to calculation of simulated radiological image patches. With this method user-controlled simulation parameters allow the generation of simulated radiological breast image patches that have realistic appearance with predefined statistical and physical properties. The approach adopted allows simulation of a wide range of realistic parenchymal patterns making patches from the simulated images ideal for use in VCTs based on the multiple alternative-forced choice (AFC) paradigm. In AFC studies, a series of target or signal detection experiments using image segments or patches (~30mm x 30mm) are presented to observers (Burgess 1995, 1999).

To test the above assertion, patches from simulated images of the breast model have been evaluated for visual realism by an ROC-based observer study with expert breast readers, and quantitatively evaluated by using fractal and power spectrum analysis.

2. Materials and methods

The breast models are produced using a biologically inspired approach whereby features and structures are extracted from DBT image planes of real breasts and used to synthesize 3D breast structure. A collection of 100 asymptomatic clinical cases acquired using a Hologic Selenia Dimensions system (Hologic, Bedford, Massachusetts, USA) rich in glandular tissue were chosen from the TOMMY trial database (Gilbert *et al* 2015) for the extraction of anatomical features. The patient cases contained both digital mammography (DM) and DBT images of the same breast in two views (cranio-caudal (CC) and (medio-lateral oblique (MLO)). The compressed breast thicknesses of the selected cases ranged between 25mm and 80mm.

Our novel simulation process starts by generating a high resolution empty breast volume surrounded by a skin layer which is then filled with voxels labelled with different tissue classes based on tissue structures extracted from DBT images. The OPTIMAM image simulation framework was then used to simulate 2D and DBT X-ray images of the breast model (Elangovan *et al* 2014, 2016). Finally, the radiological images were diced into segments as required for AFC studies. Details of each step in the process are described below.

2.1. Breast outline and skin layers

Single pixel-width outlines of breast skin contours were extracted by simple thresholding from DBT patient images to define an initial overall 3D shape. The outlines were then re-sampled to have an isotropic voxel resolution of 100µm, and this voxel size was used for all tissue components within the breast model. Each collection of plane-by-plane contours from a particular breast was then converted into a mesh-based surface representation. A 1.5mm thick tissue layer was added to the ‘skin’ surfaces of the breast outline to represent the skin layer (Ulger *et al* 2003, Huang *et al* 2008). All interior voxels of the volume were initially labelled as adipose tissue, this being subsequently referred to as the uniform adipose volume.

2.2. Simulation of glandular tissue

Segments of glandular tissue were extracted from DBT images using an adaptive seeded region growing method (Gonzalez and Woods 2007). In order to select good quality DBT data for extraction of glandular tissue, 2D images produced by summation of the DBT planes and clinical 2D digital

¹ The OPTIMAM project is a 5 year research programme aimed at defining optimal parameters and methods for mammography to improve cancer detection using the Virtual Clinical Trials approach. See <https://medphys.royalsurrey.nhs.uk/nccpm/?s=optimam-main>

mammography counterparts were visually inspected and compared. When these were judged as being visually similar then the DBT data were considered to contain minimum reconstruction artefacts and were chosen as candidates for glandular tissue extraction. Specific glandular tissue rich regions were targeted in the central region of the breast for glandular tissue extraction as explained below.

Within a bounding box that defines a glandular tissue region in the chosen DBT volume, all voxels occupying the highest intensity value were chosen as seed points for multiple 3D region growing. Subsequently, a region growing method was used to iteratively aggregate voxels to each of these regions when the difference between the mean intensity of the growing region and a neighbouring voxel was less than an empirically derived threshold value. The mean of the cluster was then updated at each iteration. This process evolved and the regions steadily grew in size until a termination condition (overall region size $> 30 \times 30 \times 30 \text{mm}^3$) was reached.

Subsequently, voxels observed to be erroneously appended to the final region were removed using connected component analysis and binary morphological operations (Gonzalez and Woods 2007). The process involved discarding voxels containing less than 8 connected neighbours and then applying a binary morphological opening operation using a disk shaped structuring element (radius 3 voxels). Figures 1(a) and 1(b) illustrate the results of glandular tissue fragment segmentation process. This process was repeated to create a database of 40 volumetric glandular fragments. Randomly selected glandular tissue fragments from the database were then combined together to produce a 3D glandular tissue matrix and placed inside the uniform adipose volume.

This approach of generating the glandular tissue matrix was developed from a breast texture synthesis method proposed by Bochud *et al* (1999) and Roland and Strickland (1997), whereby breast texture was created by superimposing Gaussian blobs at locations chosen by a random uniform distribution. The process starts by placing a random number of points within the uniform adipose breast volume. Each point serves as the insertion location for a randomly chosen glandular fragment. The number of glandular tissue fragments was augmented by rotating and scaling the fragment volumes at random orientations and sizes respectively.

Different texture patterns can be simulated by changing the fragment scaling a . A large value of a will result a lumpy texture and a small value of a will result in a finer texture. Whenever there was an intersection in the placement of glandular fragments in the breast volume, the fragments were aggregated into one large fragment. This process was repeated by adding more glandular fragments to the glandular tissue matrix until the desired user-defined glandularity and visual complexity was achieved. Figure 1(c) illustrates how the complexity of the breast texture changes as more glandular fragments are added to the glandular tissue matrix. The placement of glandular fragments was confined to the central region of the uniform adipose breast volume, allowing for a pure adipose tissue layer just below the skin layer as typically found in a real breast (Geddes 2009).

The power spectrum of breast texture produced using such a method was calculated by Bochud *et al* (1999). These authors observed that for large glandular fragments and a low value of the number of fragments K , the shape of individual fragments dominated the breast texture and the power spectrum of the radiological images of the breast model. In contrast, for smaller glandular fragments and large K , the power spectrum was dominated by large aggregations of glandular tissue fragments and not by the individual component fragments. Hence, the size of the extracted glandular fragments was limited to $\leq 30 \times 30 \times 30 \text{mm}^3$ to avoid dominant structures in the final breast volume and in the simulated radiological images.

Figure 1 (a) Segmentation of glandular tissue; (b) the 2D projection of a segmented glandular fragment; and (c) simulation of glandular tissue matrix by adding glandular tissue fragments. The images show the 2D projection of the 3D breast volume for different numbers K of glandular fragments.

Figure 2 (a) DBT plane with selected in-focus landmarks; (b) segments of the 3D spline interpolated wireframe overlaid on a single DBT plane; and (c) the wireframe dilated into blood vessels and Cooper's ligaments.

2.3. Simulation of Cooper's ligaments and blood vessels

Prominent linear structures visible in a radiological image of a female breast include Cooper's ligaments and blood vessels. To represent these in the simulated volume, a collection of 3D curves representing Cooper's ligaments and blood vessels was first produced as follows: landmarks were manually placed where such linear structures (blood vessels or Cooper's ligaments) appeared in-focus when scrolling through a stack of DBT planes (Figure 2(a)). These landmark points were then connected by a series of 3D splines to create 3D wireframe representations of these linear structures (Figure 2(b)).

The single voxel-width wireframes were then dilated (Figure 2(c)) to 3-4mm diameter for the blood vessel network, and to 1-2mm in diameter for the Cooper's ligament network. The dilation diameter was chosen depending on the linear structure characteristics of the particular DBT images from which the wireframes were extracted. Finally, the dilated linear components were integrated into the glandular/adipose breast volume containing the glandular tissue matrix by voxel replacement. Where these components coincided with voxels labelled as glandular or adipose, the linear component replaced any prior glandular or adipose voxel label.

3. Validation Methodology

Simulated radiological images of the breast models were produced and validated against real images both quantitatively and qualitatively as described below.

3.1. Synthetic image acquisition

Radiological projections of the breast model were generated by using the validated radiological image simulation chain described in Elangovan *et al* (2014). The radiological images were simulated to mimic the Hologic Selenia Dimensions 3D system. The Hologic system produces a 2D image (pixel pitch: 70 μ m) and 15 DBT projections (pixel pitch: 140 μ m; angular range: +7.5° to -7.5°) when operating in ‘combo’ mode. The system is equipped with an amorphous selenium detector of size 29 $\times$ 24cm². Primary projections of the breast model were produced using a ray tracing tool (Siddon 1985) for a predefined set of technical factors including the X-ray spectrum. The X-ray spectrum was derived from the spectral model of Boone *et al* (1997) and was scaled to represent the incident air kerma at the upper surface of the breast corresponding to the particular mean glandular dose (MGD) being simulated. The attenuation values for different tissue components of the breast model were derived from the elemental compositions of the adipose tissue, glandular tissue from Hammerstein *et al* (1979), whereas, the composition of Cooper’s ligaments and blood vessels were assumed to be that of adult skeletal muscle (ICRU 1992, Ullman *et al* 2003). An offset was added to the images to simulate the contrast reduction effect due to scatter. This offset was derived from a scatter-to-primary ratio look-up table constructed from Monte Carlo simulation data for standard breast models (Diaz *et al* 2014). We used the methods of Mackenzie *et al* (2012, 2014) to add intrinsic system blur and include the effect of three major noise sources (system, electronic and quantum). Finally, the intensity values were converted into pixel values from the known signal transfer properties of the detector. In addition, X-ray tube motion blur was also included in the simulations by elongating the focal spot in the direction of the tube motion (Elangovan *et al* 2014). A conventional 2D-mammography projection image and 15 DBT projection images were produced. The 2D image was then processed using the manufacturer’s image processing tool (Hologic LORAD FFDM Selenia V5.0). DBT planes were produced using the manufacturer’s image reconstruction tool which employs a filtered back-projection technique. Figure 3 shows examples of patches from processed 2D images and reconstructed DBT planes produced using the image simulation chain.

Figure 3 2D and corresponding DBT images simulated using the breast models with (a) a fine texture pattern ($a=0.5$) (b) a clustered background texture ($a=1.0$) (c) a lumpy background ($a=1.5$). a is the scaling factor and K is the number of glandular fragments, set to 200 for all three models.

For validation purposes, five different breast thicknesses were simulated: 30mm, 40mm, 50mm, 60mm and 70mm. At each thickness, three glandular patterns (as shown in Figure 3) were simulated resulting in 15 breast models altogether. The simulated percentage glandularity of each breast model was varied with breast thickness according to density measurements performed on a set of real images using the Volpara breast density measurement tool (Volpara Health Technologies Limited, Wellington, New Zealand) (Highnam *et al* 2010) (Table 1). The simulated glandularity for the breast models was chosen from the Volpara data (volumetric breast density) derived from the TOMMY trial database images. In this work we selected ~300 images from the TOMMY trial database which predominantly contained images rich in dense glandular tissue, chosen for the purposes of glandular tissue extraction. These were grouped into thickness class intervals as shown in Table 1, and then the mean glandularity, using Volpara, was calculated for each thickness class. These glandularities are higher than those found in a general screening population (Youn *et al* 2016, Ng and Lau 2015) due to this image selection process. However, the simulation allows generation of breast models with any combination of thickness and glandularity, so that the combinations shown are merely illustrative of what can be achieved with this approach. We have validated our model to simulate central areas of the breast with relatively high glandularity. The technical image acquisition factors (Table 1) for the radiological image simulation step were chosen from a look-up table replicating the automatic exposure control settings of a Hologic Selenia Dimensions 3D system measured using PMMA (polymethyl methacrylate) phantom blocks (Elangovan *et al* 2014).

Once the above database of anatomical features (glandular segments, Cooper’s ligaments and blood vessels) had been constructed, the above approach requires, on average, 10-20 mins to simulate a 6cm thick volume with an average glandularity (excluding skin layer) of 20% by volume and corresponding 2D and DBT images. This can then be diced into a large number of unique image segments (~30mm x 30mm) with predefined statistical and physical properties. These can be produced over a relatively short time interval (500 radiological image segments per day) for use in AFC virtual clinical trials. This simulation time is for a single MATLAB session running on a quad core Linux machine with 3GB RAM. The simulation time increases with increased glandularity and breast thickness, according to the number of glandular fragments required to achieve the user defined density.

Table 1 Summary of simulation parameters.

Breast model thickness (mm)	Volumetric breast density (%)	2D		DBT	
		kVp	Target/Filter	kVp	Target/Filter
30	71%	27	W/Rh	29	W/Al
40	46%	28	W/Rh	30	W/Al
50	30%	30	W/Rh	32	W/Al
60	20%	32	W/Rh	34	W/Al
70	10%	31	W/Ag	36	W/Al

3.2. Clinical image selection

For validation, we selected a set of real 2D and DBT images. The clinical cases were randomly chosen from the TOMMY trial database (Gilbert *et al* 2015), comprising images from women recalled after routine screening and women recalled after being screened due to a family history of breast cancer. The clinical cases chosen for the study included 50 asymptomatic cases (2D images and associated DBT datasets) that had been labelled as free of lesions and free of architectural distortion acquired using a Hologic Selenia Dimensions system. The images included both CC and MLO projections. There was no overlap between the cases used for model development and validation. The breast thicknesses of the cases ranged between 26 mm and 83 mm. For validation, ROIs were manually extracted from the images and were limited to the central region of the breast where thickness would be expected to be constant.

3.3. Fractal dimension

Fractal dimension was used to compare the complexity of simulated images with that of real images following published methods (Caldwell *et al* 1990; Bliznakova *et al* 2010). The fractal dimension of an image surface can be calculated using equation (1)

$$A(\varepsilon) = \lambda \varepsilon^{2-D} \quad (1)$$

where $A(\varepsilon)$ is the area of the image surface, λ is a scaling factor and D is the surface fractal dimension. Each pixel in the image was assumed to be part of a surface intensity map with pixel pitch ε , and height equal to the pixel intensity value $I(x,y)$. The area $A(\varepsilon)$ of the surface elevation was calculated as the sum of the area of the pixels and the exposed lateral sides of the pixel intensity “heights” as shown in equation (2).

$$A(\varepsilon) = \sum_{x,y} \varepsilon^2 + \sum_{x,y} \varepsilon \{|I(x,y) - I(x+1,y)| + |I(x,y) - I(x,y+1)|\} \quad (2)$$

The measurements were repeated for different pixel dimensions ε obtained by subsampling the image at different scales, ε , and averaging the intensity values within ε . Finally, the fractal dimension of the image ($2-D$) was computed from the slope D of $\log(A(\varepsilon))$ versus $\log(\varepsilon)$.

3.4. Power spectrum analysis

Power spectrum analysis was used to compare the statistical properties of real and simulated images. The anatomical structures seen in mammograms are known to be described by a power-law spectrum (Burgess 1999, Cockmartin *et al* 2013, Hill *et al* 2013), of the form shown in equation (3), where β is the power-law exponent which reflects breast tissue complexity; α is the power spectrum magnitude in units of area; and f is the spatial frequency.

$$P(f) = \alpha / f^\beta \quad (3)$$

Power spectrum analysis was performed on unprocessed 2D and reconstructed DBT planes of the breast models and real images following published methods (Cockmartin *et al* 2013, Hill *et al* 2013). The ROI size was set to 25mm² with an overlap of 1.25mm in both horizontal and vertical directions. The power spectra were estimated by computing the squared magnitude of the Fourier transform of the ROIs extracted from the image. The ROIs were limited to the central region of the breast. The power spectra ensembles of ROIs were normalized by dividing by the mean pixel value of the ROIs. A Hanning window was applied to the ROIs prior to the Fourier transform step to reduce spectral leakage. The radially average power spectra frequencies in the range 0.2-0.7mm⁻¹ were used in the analysis, as this range is dominated by spatial texture properties of the breast. Thus power-law parameters were estimated by applying a linear fit to the log transformed data in the frequency range of 0.2-0.7mm⁻¹. The power spectra for DBT images were calculated in the same manner as the 2D images except the image pixels were not normalised since they are independent of the dose applied (Cockmartin *et al* 2013, Hill *et al* 2013).

The incident air kerma at the top of the breast for the simulated images was set to be similar to that of the real images used in the study. This permitted a fair comparison since the same amount of quantum noise will be present in both real and simulated images.

Student t-tests were used to test the statistical significance of any difference between fractal dimension and power spectrum values calculated from the real and simulated images.

3.5. ROC Observer study

Visual realism was assessed using blinded observer studies in which experienced breast readers rated the appearance of a mixture of simulated and real image patches of size 30mm x 30mm. The dataset contained 35 real and 35 simulated 2D and DBT patches. Prior to random selection, the simulated images were visually screened and image patches that appeared too geometric due to the presence of overlapping linear structures were removed. This was a fast process performed by non-clinical staff.

Seven observers with clinical screening experience ranging between 5 and 25 years participated. All the observers had extensive experience viewing 2D mammography screening images. Two of the observers (observers 3 & 7) had significant experience viewing DBT images and the rest had a more limited experience with DBT, as this modality is not used in routine breast screening in the UK. Each observer was asked to rate the realism using a 6 point scale as shown in Figure 4. Each DBT image volume was presented as an image stack through which the observer was able to scroll. The size of the image segments was 3cmx3cm, appropriate for use in a 4AFC paradigm. The study was conducted using a dedicated workstation at low ambient room lighting (<6 lux) with no time limitations. The observers had no knowledge of the percentage of real and simulated images in the dataset. The areas under the ROC curves (AUC) for 2D and DBT were computed using the trapezoidal rule and the standard error was calculated using the method described in Hanley and McNeil (1982).

Figure 4 Graphical user interface used for the study.

3. Results and Discussion

The mean fractal dimensions of real and simulated images are shown in Table 2. There was no significant difference between these fractal dimensions ($p>0.05$) for both 2D and DBT images.

Table 2 Fractal dimension of real and simulated images. (The errors are 2 standard errors of the mean.)

	Real images	Synthetic images	P-value
2D	2.77 ± 0.03	2.72 ± 0.01	0.08
DBT	2.82 ± 0.04	2.75 ± 0.01	0.09

Figures 5 and 6 show the average power spectrum plots of simulated and real images for 2D and DBT. The dotted lines represent 95% confidence intervals. Table 3 shows the alpha and beta values calculated for real and simulated images. The beta value of real images used in this study ($\beta \approx 3$) is consistent with previously published work (Burgess 1999, Cockmartin *et al* 2013, Hill *et al* 2013). The average beta values of simulated images closely match those of real images for both 2D (<7% difference) and DBT (<4% difference). There was no significant difference between the beta values ($p>0.05$) for the real and simulated 2D images ($p=0.58$) and DBT images ($p=0.59$). There is more variability in the fractal dimension and power spectrum values of the real images compared to the simulated images. This is expected because of the limited range of texture patterns that can be currently simulated in the breast model. Overall, the results of the above analyses indicate that the statistical properties of radiological images produced from the breast tissue model are in excellent agreement with those of real images.

Figure 5 Normalized power spectrum curves for 2D images. The vertical dashed lines represent the range of spatial frequencies attributable to breast texture rather than other factors such as quantum noise. The solid lines represent the average value and the dotted lines represent the 5th and 95th percentiles.

Figure 6 Power spectrum curves for DBT images. The vertical dashed lines represent the range of spatial frequencies attributable to breast texture rather than other factors such as quantum noise. The dotted lines represent 5th and 95th percentile. The solid lines represents the average value and the dotted lines represent the 5th and 95th percentiles.

Table 3 Power law characteristics of real and simulated images.

	Real images		Simulated images	
	β (min,max,SE)	α (min,max,SE)	β (min,max,SE)	α (min,max,SE)
2D	3.01 (2.02;3.90;0.320)	0.008 (0.005;0.017;0.002)	3.19 (3.07;3.31;0.070)	0.006 (0.005;0.007;0.001)
DBT	3.10 (2.70;3.33;0.170)	0.013 (0.010;0.017;0.002)	3.21 (3.01;3.42;0.110)	0.012 (0.001;0.014;0.001)

The ROC plots of the observer study are shown in Figures 7 and 8. The diagonal solid line represents 50-50 chance. The mean AUC was 0.51 ± 0.06 for 2D images and 0.54 ± 0.09 for DBT images, with the errors expressed as the standard errors of the mean AUC. The proximity of both ROC curves to the mean chance line indicates that the observers had difficulty in distinguishing real images from simulated images. It is noted that observer 3 had a higher AUC than other observers for both 2D and DBT images. This observer was a consultant radiologist with more than 20 years' experience in breast imaging. In contrast, observer number 7, who had similar experience as observer 3 in breast imaging and extensive experience reading DBT images performed no better than less experienced observers. Following the study, observer 3 was questioned on rating certain images as simulated. It transpired this rating was applied when linear structures in the images appeared 'too geometric' or 'too perfect'. This can be attributed to the 3D spline interpolation technique used for generating high frequency structures from the DBT in-focus landmarks which might produce such a geometric effect in some cases. It should also be noted that observer 3 had prior knowledge about the breast model development, which may have influenced their ratings. We have previously seen a similar effect during validation of our lesion simulation model (Rashidnasab *et al* 2013a) where prior knowledge of the methodology tended to produce higher AUC values. In light of this, the average AUC was recomputed after excluding observer 3 and found to be 0.49 ± 0.01 for 2D images and 0.51 ± 0.01 for DBT images which shows a ~4% relative difference. Overall, the results suggest that the patches from mammographic/DBT images of the breast models produced by simulating image acquisition of the Hologic Dimensions 3D system may be considered to be sufficiently indistinguishable from patches from real screening images acquired using the same system to be used in AFC VCT studies.

Figure 7 ROC curves for 2D images.

Figure 8 ROC curves for DBT images.

Table 4 shows individual ROC values and confidence intervals for all the observers along with years of experience in breast imaging. Spearman's rank correlation coefficient was used to test the relationship between observer performance and screening experience. The association between the two variables was not statistically significant for both 2D ($p=0.31$) and DBT images ($p=0.67$).

Table 4 AUC for each observer. The errors represent one standard error of the mean.

Observer	Years of experience	2D	DBT
1	18	0.50±.08	0.49±.07
2	10	0.48±.09	0.52±.04
3	20	0.64±.10	0.73±.15
4	5	0.44±.11	0.55±.04
5	14	0.51±.07	0.49±.08
6	10	0.53±.05	0.53±.03
7	20	0.48±.09	0.48±.08
Mean	13.5±2.36	0.51±.03	0.54±.04

This work represents the final step in developing a set of validated baseline tools for VCTs using MAFC methodology. Thus the OPTIMAM VCT Toolbox is now able to create simulated anatomy/tissue into which validated simulated pathology can be inserted for creating realistic radiological image patches. Of course this method is not without its limitations. The glandular fragments used for constructing the breast model were extracted from DBT images contained in the TOMMY Trial database acquired using Hologic systems. This might make the breast model somewhat manufacturer-specific. According to Cockmartin *et al* (2013), Siemens reconstruction filters have a very strong influence on the power spectrum curves of DBT images compared to Hologic reconstruction filters. This was apparent from the differences between the power spectrum curves of DBT projection and reconstructed planes of both systems. One can speculate that the glandular fragments extracted from different systems might affect the breast model in different ways. Further simulation and validation studies need to be conducted with other systems such as Siemens and GE to explore any such potential issues. Reconstructed DBT volumes from which glandular tissue fragments are extracted suffer from poor z-resolution and also contain additional blur due to limited angle reconstruction. Hence, a set of morphological operations including erosion were used to remove small erroneous regions from the segmented glandular fragments. However, a small amount of additional blur in the images calculated from the breast models may be present.

Further improvements can be made to the Cooper's ligament and blood vessel network simulation methods. Currently, Cooper's ligaments are simulated as 1mm cylinders. In reality, Cooper's ligaments are far more complex structures made up of sheets of connective tissue that help maintain the structural integrity of the breast (Graff 2016, Bakic 2011). Similarly, blood vessel simulation is currently limited to large vessels and simulated as 3mm cylinders. In reality, the breast vascular system is a complex network of axillary (large), internal mammary (medium) and intercostal (small) arteries and veins (van Deventer 2005). More accurate simulations could be achieved by simulating these structures using voxels smaller than 100 micron. However, with this approach, computational time and memory requirements will limit the practicalities of using simulations. In the current study, the size of Cooper's ligaments and blood vessels for the simulation were estimated from clinical 2D and DBT images. The data on the size of the linear structures reported in the literature are variable (Graff 2016, Bakic *et al* 2011, Bliznakova *et al* 2003). In the future more accurate estimates of the sizes can be used in the simulations to improve the realism. This would allow us to randomly choose clinically-realistic image patches from the breast models for observer studies without the need for prior visual selection.

In reality, breast tissue exhibits directionality, with a preferred orientation towards the nipple (Reiser *et al* 2012). However, the proposed method of glandular tissue simulation assumes that the glandular tissue distribution within the breast is fairly uniform and exhibits no directionality. Although lack of directionality may not significantly affect the results of the power spectrum analysis as shown by Reiser *et al* (2012), this will affect the visual realism of images. For small patches, this appears to be insignificant; however in larger images including full size breast phantoms the effect will be significant. The distribution of glandular tissue in real images can range from fluffy blobs to sparse lines (Miller and Astley 1991). At the moment different texture patterns are simulated by only changing the coarseness of the glandular tissue fragments. As a result, images calculated from the breast model represent only a limited range of patterns typically found in real world clinical breast data. This is apparent from the narrow confidence intervals of fractal dimension and power spectrum results compared to a much wider range exhibited in clinical data. The primary motivation to date for this work has been for application in the AFC paradigm, so that large scale tissue deformations associated with e.g. ductal anatomy, glandular tissue orientation and distribution are not significant. The next steps will focus on incorporating these aspects to refine the current model to create full-size realistic mammograms for VCTs involving search and localization. Currently, the image patches for the 4AFC studies are extracted from the central region of the breast models where thickness is expected to be constant, and hence a spatially constant thickness dependent scatter offset is added to the images during simulation. Although breast glandularity will have some effect on the scatter, the predominant effect will be from the change of breast thickness (Diaz *et al* 2014). If the models are to be extended to a full breast phantom, the changes in scatter across the breast should be taken into account. This could be achieved using Monte Carlo simulations to produce scatter map for each phantom, or using an extended table of scatter to primary ratio data for a range of standard breast models. This should be followed by an extended study validating the suitability of images calculated from phantoms for replacing clinical images in diagnostic studies.

4. Conclusions

The use of clinically realistic simulated images is key to ensuring the validity of VCT results. The proposed method can rapidly produce a multiplicity of different breast appearance models as needed for conducting 4AFC VCTs. In this work, a fast method for simulating breast tissue using a biologically inspired approach has been developed. The quantitative validation has demonstrated strong agreement between fractal dimension values and power spectrum curves obtained for real and simulated images. Validation of visual realism was successfully undertaken using an ROC study for a blinded comparison to real images using expert breast screening readers. Our results suggest that the images are indistinguishable from screening images for patches of 30mm x 30mm. Taken together with previous published work, the OPTIMAM VCT toolbox now forms a set of end-to-end VCT simulation tools for use in MAFC trials that have been validated for quantitative correctness and visual realism.

Acknowledgements

This work is part of the OPTIMAM2 project funded by Cancer Research UK (grant number: C30682/A17321). We are grateful to Hologic for the provision of the image processing and reconstruction software. We thank radiologists and radiographers from St Georges Hospital, London for participating in the observer study. We thank Professor Fiona Gilbert at Addenbrooke’s Hospital Cambridge for providing access to images in the TOMMY trial database.

References

Bakic P R, Albert M, Brzakovic D and Maidment A D A 2002a Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic mammogram texture *Med. Phys.* **29**, 2140–2151.
Bakic P R, Albert M, Brzakovic D and Maidment A D A 2002b Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation *Med. Phys.* **29**, 2131–2139.

- 1
- 2
- 3 Bakic P R, Zhang C and Maidment A D A 2011 Development and characterization of an
- 4 anthropomorphic breast software phantom based upon region-growing algorithm *Med. Phys.* **38**,
- 5 3165-3176.
- 6 Bakic P, Pokrajac D, De Caro R, Maidment A 2014 Realistic Simulation of Breast Tissue
- 7 Microstructure in Software Anthropomorphic Phantoms *Proc IWDMM '14*. 8539, 348-355.
- 8 Bliznakova K, Bliznakov Z, Bravou V, Kolitsi Z and Pallikarakis N 2003 A three-dimensional breast
- 9 software phantom for mammography simulation *Phys. Med. Biol.* **48**, 3699-3719.
- 10 Bliznakova K, Suryanarayanan S, Karellas A and Pallikarakis N 2010 Evaluation of an improved
- 11 algorithm for producing realistic 3D breast software phantoms: Application for mammography
- 12 *Med. Phys.* **37**, 5604-5617.
- 13 Bliznakova K, Sechopoulos I, Buliev I and Pallikarakis N 2012 BreastSimulator: A Software Platform
- 14 for Breast X-ray Imaging Research *J. Biomed. Graph. Comp.* **2**, 1-14.
- 15 Boone J M, Fewell T R and Jennings R J 1997 Molybdenum, Rhodium and Tungsten anode spectral
- 16 models using interpolating polynomials with application to mammography *Med. Phys.* **24**, 1863-
- 17 1874.
- 18 Bochud F O, Abbey C K and Eckstein M P 1999 Statistical texture synthesis of mammographic
- 19 images with clustered lumpy backgrounds *Opt Express* **4**, 33-42.
- 20 Burgess A E 1995 Comparison of receiver operating characteristic and forced choice observer
- 21 performance measurement method *Med. Phys.* **22** 643-55
- 22 Burgess A E 1999 Mammographic structure: Data preparation and spatial statistics analysis *Proc.*
- 23 *SPIE 3661*, 642-653.
- 24 Castella C, Kinkel K, Descombes F, Eckstein M P, Sottas P E, Verdun F R and Bochud F O. 2008
- 25 Mammographic texture synthesis: Second-generation clustered lumpy backgrounds using a genetic
- 26 algorithm *Opt Express* **16**, 7595-7607.
- 27 Caldwell C, Stapleton S, Holdsworth D, Jong R, Weiser W, Cooke G and Yaffe M 1990
- 28 Characterisation of mammographic parenchymal pattern by fractal dimension *Phys. Med. Biol.* **35**,
- 29 235-347.
- 30 Cockmartin L, Bosmans H and Marshall N W 2013 Comparative power law analysis of structured
- 31 breast phantom and patient images in digital mammography and breast tomosynthesis *Med. Phys.*
- 32 **40**, 081920.
- 33 Dance D R, Strudley C J, Young K C, Oduko J M, Whelehan P J and Mungutroy E H L 2014
- 34 Comparison of breast doses for digital tomosynthesis estimated from patient exposures and using
- 35 PMMA breast phantoms *Phys. Med. Biol.* **59**, 4275-93.
- 36 Diaz O, Dance D R, Young K C, Elangovan P, Bakic P R and Wells K 2014 A fast scatter field
- 37 estimator for digital breast tomosynthesis *Phys. Med. Biol.* **59**, 2375-2390.
- 38 Diekmann F and Bick U 2007 Tomosynthesis and contrast-enhanced digital mammography: recent
- 39 advances in digital mammography *Eur. Radiol.* **17**, 3086-3092.
- 40 Elangovan P, Warren L M, Mackenzie A, Rashidnasab A, Diaz O, Dance D R, Young K C, Bosmans
- 41 H, Strudley C J and Wells K 2014 Development and validation of a modelling framework for
- 42 simulating 2D-mammography and breast tomosynthesis *Phys. Med. Biol.* **59**, 4275-4293.
- 43 Elangovan P, Hadjipanteli A, Mackenzie A, Dance DR, Young KC and Wells K 2016 OPTIMAM
- 44 image simulation toolbox – recent developments and ongoing studies *Proc. IWDMM 9699*, 668-675.
- 45 Geddes D T 2009 Ultrasound imaging of the lactating breast: methodology and application *Intl.*
- 46 *Breastfeeding Jnl* **4**.
- 47 Gilbert F J *et al* 2015 The TOMMY trial: a comparison of TOMosynthesis with digital
- 48 MammographY in the UK NHS Breast Screening Programme – a multicentre retrospective reading
- 49 study comparing the diagnostic performance of digital breast tomosynthesis and digital
- 50 mammography with digital mammography alone *Health Technology Assessment* **19**.
- 51 Gilbert F J, Tucker L and Young K C 2016 Digital Breast Tomosynthesis (DBT): A review of the
- 52 evidence for use as a screening tool *Clinical Radiology* **71**, 141-50.
- 53 Gong X, Glick S J, Liu B, Vedula A A and Thacker S 2006 A computer simulation study comparing
- 54 lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT
- 55 breast imaging *Med. Phys.* **33**, 1041-1052.
- 56 Gonzalez R C and Woods R E 2007 Digital Image Processing 3rd edn. Pearson New Jersey, NY.
- 57
- 58
- 59
- 60

- Graff C G 2016 A new open-source multi-modality digital breast phantom *Proc SPIE*. **9783**, 9783091-10.
- Hanley J A and McNeil B J 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve *Radiology* **143** 29–36.
- Hammerstein G R, Miller D W, White D R, Masterson M E, Woodard H Q and Laughlin J S 1979 Absorbed radiation dose in mammography *Radiology* **130** 485–91 (1979).
- Heine J J, Deans S R, Velthuizen R P and Clarke L P 1999 On the statistical nature of mammograms *Med. Phys.* **26**, 2254-2265.
- Hevie M A, Chang J T, Hendrick R E and Banerjee M 2014 Reduction in late-stage breast cancer incidence in the mammography era: Implications for overdiagnosis of invasive cancer. *Cancer* **120**, 2649-2656.
- Hill M L, Mainprize J G, Carton A K, Saab-Puong S, Iordache R, Muller S, Jong R A, Dromain C and Yaffe M J 2013 Anatomical noise in contrast-enhanced digital mammography. Part I. Single-energy imaging *Med. Phys.* **40**, 051910.
- Hoeschen C, Fill U, Zankl M, Panzer W, Regulla D and Dohring W 2005 A high resolution voxel phantom of the breast for dose calculations in mammography *Radiat. Prot. Dosim.* **114**, 406–409.
- Huang S Y, Boone J M, Yang K, Kwan A L and Packard N J 2008 The effect of skin thickness determined using breast CT on mammographic dosimetry *Med. Phys.* **35**, 1199-206.
- ICRU 1992 Photon, electron, proton and neutron interaction data for body tissues *Tech. Rep. International Commission on Radiation Units and Measurements* **46**.
- Ikejimba L, Lo J Y, Chen Y, Oberhofer, N, Kiarashi, N, Samei, E 2016 A quantitative metrology for performance characterization of five breast tomosynthesis systems based on an anthropomorphic phantom *Med. Phys.* **43**, 1627.
- Karim-Kos H E, de Vries E, Soerjomataram I, Lemmens V, Siesling S and Coebergh J W 2008 Recent trends of cancer in Europe: A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s *European Journal of Cancer* **44**, 1345-1389.
- Kiarashi N, Nolte L W, Lo J Y, Segars W P, Ghate S V and Samei E 2015 The impact of breast structure on lesion detection in breast tomosynthesis. *Proc. SPIE* **9412**, 941229.
- Li C M, Segars W P, Tourassi G D, Boone J M and Dobbins J T 2009 Methodology for generating a 3D computerized breast phantom from empirical data *Med. Phys.* **36**, 3122–3131.
- Mackenzie A, Dance D R, Workman A, Yip M, Wells K and Young K C 2012 Development and validation of a method for converting images to appear with noise and sharpness characteristics of a different detector and X-ray system *Med. Phys.* **39**, 2721-2734.
- Mackenzie A, Dance D R, Diaz O and Young K C 2014 Image simulation and a model of noise power spectra across a range of mammographic beam qualities *Med. Phys.* **41**, 12901-14.
- Maidment A D A 2014 Virtual clinical trials for the assessment of novel breast screening modalities. *Proc IWDW '14*. **8539**, 1-8.
- Ng K and Lau S 2015 Vision 20/20: Mammographic breast density and its clinical applications *Med. Phys.* **42**, 7059–7077.
- Nooshin K, Nolte L W, Lo J Y, Segars W P, Ghate S V, Solomon J B, Samei E 2016 Impact of breast structure on lesion detection in breast tomosynthesis, a simulation study *J. Med. Imag.* **3**, 035504.
- Niklason L T *et al* 1997 Digital tomosynthesis in breast imaging *Radiology* **205**, 399–406 (1997).
- O'Connor J M, Das M, Didier C, Mah'D M and Glick S J, 2008 Comparison of two methods to develop breast models for simulation of breast tomosynthesis and CT *Proc. IWDW '08*, 417-425.
- Miller P and Astley S 1991 Classification of breast tissue by texture analysis *Img. Vis. Comp.* **10** 277-282.
- Pokrajac D D, Maidment A D, Bakic P R 2012 Optimized generation of high resolution breast anthropomorphic software phantoms *Med. Phys.* **39**, 2290-2302.
- Rashidnasab A, Elangovan P, Diaz O, Mackenzie A, Young K C, Dance D R and Wells K 2013a Simulation of 3D DLA masses in digital breast tomosynthesis *Proc. SPIE* **8668**, 86680Y.
- Rashidnasab A, Elangovan P, Yip M, Young K C, Dance D R and Wells K 2013b Simulation and assessment of realistic breast lesions using fractal growth models *Phys. Med. Biol.* **15**, 5613-5626.
- Reiser I, Lau B A, Nishikawa R M and Bakic P R 2012 A directional small-scale tissue model for an anthropomorphic breast phantom. *Proc IWDW '12*. **7361**, 141-148.

- 1
2
3 Rolland J P and Strickland R N 1997 An approach to the synthesis of biological tissue *Opt. Express* **1**,
4 414–423.
- 5 Ullman G, Sandborg U M, Hunt R, Dance D R and Carlsson A G 2003 Implementation of pathologies
6 in the Monte Carlo model in chest and breast imaging. *Tech. Rep. Institutionen för radiologi,*
7 *Universitetet i Linköping* **94**.
- 8 Shaheen E, Zanca F, Sisini F, Zhang G, Jacobs J and Bosmans H 2010 Simulation of 3D objects into
9 breast tomosynthesis images *Radiat. Prot. Dosim.* **139**, 108–112.
- 10 Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional CT array *Med.*
11 *Phys.* **12**, 252–255.
- 12 Ulger H, Erdogan N, Kumanlioglu S and Unue E 2003 Effect of age, breast size, menopausal and
13 hormonal status on mammographic skin thickness *Skin Res. Technol.* **9**, 284–289.
- 14 van Deventer P V 2005 The blood supply to the nipple-areola complex of the human mammary gland
15 *Aes. Plast. Surg.* **28**, 393–398.
- 16 Youn I, Choi S, Kook A H and Choi Y J 2016 Mammographic breast density evaluation in Korean
17 women using fully automated volumetric assessment *J. Korean Med Sci.* **31**, 457–462.
- 18 Young S, Bakic P R, Myers K J, Jennings R J and Park S 2013 A virtual trial framework for
19 quantifying the detectability of masses in breast tomosynthesis projection data *Med. Phys.* **40**,
20 051914:1–051914:15.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60