Time-Domain Microwave Radar Applied to Breast Imaging: Measurement Reliability in a Clinical Setting

Emily Porter[*]**, Adam Santorelli, and Milica Popović**

Abstract—This work presents an evaluation of the measurement challenges in clinical testing of our microwave breast cancer screening system. The time-domain radar system contains a multistatic 16-antenna hemi-spherical array operating in the 2–4 GHz frequency range. We investigate, for the first time with such a system in clinical trials, the repeatability of measurements and its effect on image reconstruction. We record vertical and horizontal measurement uncertainties under different scenarios and verify using previously introduced compensation methods that they can be successfully reduced to an acceptable level from the standpoint of image reconstruction. We also examine how placement of an immersion medium can affect collected breast scan data. Finally, we probe the repeatability and consistency of measurements with patients. With the goal of confirming the feasibility of frequent breast health monitoring, with our system, we obtain a total of 342 breast scans collected over 57 patient visits to determine how much scan data varies when there are no changes in between scans, and how much it varies when the patient is repositioned in the system. We confirm that, by taking care in patient positioning in the system and with respect to the immersion medium, the measurement repeatability is high.

1. INTRODUCTION

Microwave technologies have been extensively researched for their potential in breast imaging and breast cancer detection. The current standard of breast cancer screening is X-ray mammography; however, it has several drawbacks including the use of ionizing radiation, painful scans, and it can suffer from poor sensitivity and specificity, particularly in the case of dense breasts [1, 2]. Microwave techniques aim to provide an alternative method that is cost-effective, comfortable, and without ionizing radiation. Such methods typically fall into one of three categories: tomography, holography, or radar. In tomographic methods, the goal is to reconstruct a dielectric profile of the breast tissues [3]. In microwave holography, coherent back-scattered signals acquired on apertures are used to reconstruct the target [4]. Alternatively, radar methods attempt to map regions of dielectric scattering, from which the tissue type can be inferred [3].

In the published studies on microwave breast imaging, only a few research groups report on systems that have undergone clinical trials [5–8]. Our system is the first microwave-radar time-domain system to reach the clinical trial stage. Time-domain measurements offer possible advantages over frequency-domain measurements, including the potential for faster scan times [9] and more cost-effective equipment solutions. The goal of our system is long-term frequent patient monitoring, allowing for timely observation of early-stage abnormality development. The motivation for regular monitoring is underscored by the statistics: if a breast tumor is detected when it is still localized, the 5-year survival rate is 99%, but if the cancer has metastasized the rate decreases to only 23% [1]. With this in mind, our system is now undergoing clinical trials for monitoring purposes; however, it is important that we

first study the repeatability of scans so that image differences due to tissue changes can be differentiated from differences due to the measurement setup.

The study presented here is, to our knowledge, the first investigation of measurement repeatability with microwave radar breast screening in a clinical scenario. For microwave breast cancer detection systems in general, several investigations dealing with measurement uncertainty, noise and general issues have been conducted: numerical studies that examined robustness of imaging algorithms to noise [10, 11], measurements and analysis of measurement uncertainty with a tomographic system [9], and a study of error due to sensor and cable movement [6]. Related to our system, in [12], we performed a study on homogeneous breast phantoms in which we identified sources of horizontal and vertical uncertainty, and saw how these uncertainties affected reconstructed breast images. Similarly, in [13] we analyzed how antennas affect the data and corresponding images generated from phantoms. In [14], we presented a small-scale study on four patients to observe the variation in scans performed on the same day. Here, we advance on both the studies in [12, 14] by performing a more thorough examination on increasingly complex scenarios: first with an immersion medium, and finally with scans of several patient volunteers. We wish to determine if the results found in [12] on phantoms hold for scans with patients, and to see if patient-related factors (movement between scans, breast positioning) jeopardize consistent quality image reconstruction. We show how these negative factors can be effectively mitigated, thus allowing for repeatable reconstructed breast images. We aim to determine the system reliability and repeatability, and to identify measures to optimize these factors, so that breast monitoring can become feasible.

2. PROTOTYPE DESCRIPTION

2.1. Measurement System

The measurement system is a time-domain radar setup. A short-duration pulse is produced by a generator (Picosecond Pulse Labs, Model 3600, 7.5 V, 70 ps full-width at half-maximum width) on every clock signal (Tektronix gigaBERT 1400, running at 1 MHz). A pulse-shaping circuit, made up of a passive Synthesized Broadband Reflector (SBR), a directional coupler, and an amplifier (Mini-Circuits ZVE-3W-83+, +35 dB typical gain, 2–8 GHz), reshapes the generic pulse such that its spectrum is focused in the 2–4 GHz range [15]. This frequency range is chosen as a trade-off between the high-resolution yet high attenuation at upper frequencies and the low attenuation with low resolution at lower frequencies. After the pulse is shaped, it is fed into a 16 : 2 automated switching matrix that selects each transmit-receive antenna pair in turn. The receive antenna is connected through the switching matrix to a time-equivalent sampling oscilloscope (picoTechnology, PicoScope 9201 Sampling Oscilloscope, running at 40 GSa/s). A schematic drawing of the measurement system is provided in Figure 1.

The antennas are wideband, and are designed specifically for use in bio-sensing applications [16]. The Travelling Wave Tapered and Loaded Transmission Line Antennas (TWTLTLAs) are embedded in a dielectric radome that performs two tasks: a) provides improved matching between the antenna and the breast tissues, and b) securely and precisely holds the antennas in position.

2.2. Interfacing with Patients

The system is positioned below a patient exam table such that all components are secured beneath the table, except for the radome, which protrudes through a hole in the table. The radome is the interface between the system and the patient. Photographs of the stack of measurement equipment and of the radome are given in Figure 2, along with an image of the exam table.

An immersion medium is needed to fill the space between the radome walls and the breast. This material prevents the formation of air gaps that could perturb signal propagation, and provides a lossy background so that multiple reflections between the radome and the skin are attenuated. The medium chosen for this test is ultrasound gel (relative permittivity $\varepsilon_r = 68$ at our centre frequency of 3 GHz) because it meets the electrical requirements and has the favorable mechanical properties of being able to conform well to surfaces without the mess of a liquid, it is already approved for medical use, and it is cost-effective.
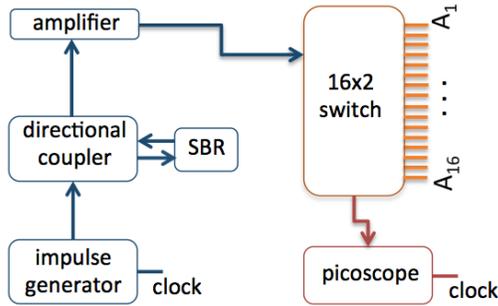
Figure 1. Schematic drawing of measurement system. The transmitting circuitry includes an impulse generator, directional coupler, SBR (synthesized broadband reflector), and amplifier; the receiving circuitry consists only of a picoscope. The switching matrix box represents the stages in between transmitting and receiving: the antennas, radome, and breast under test (figure adapted from [12]). "A_x" where $x = 1 : 16$ are the 16 antennas.
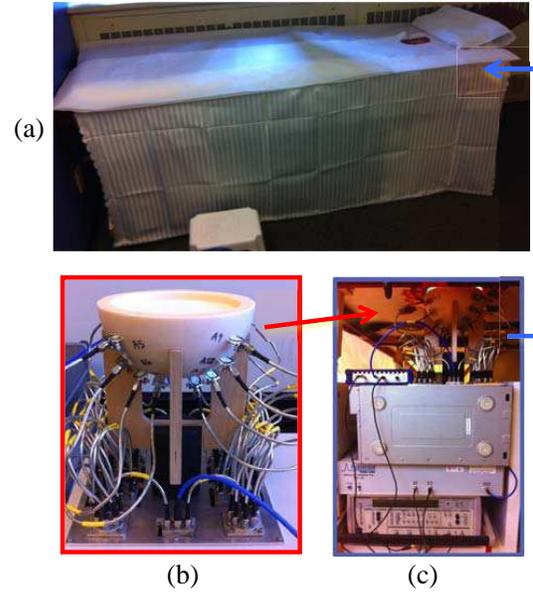
Figure 2. Photographs of (a) the patient exam table; (c) measurement equipment underneath the table; and (b), [12] a close-up of the radome and antenna array.

3. SOURCES OF MEASUREMENT UNCERTAINTY AND ERRORS

As our system has many components, there are several sources of uncertainties and errors. We separate the sources into two types: measurement uncertainty due to electronic components and devices, and errors due to the physical setup of the system interface with the patient or breast model.

The components that contribute to the measurement uncertainty are as follows. The clock jitter contributes ± 4 ns of horizontal uncertainty over five periods. The picoscope trigger specifications list less than 3.5 ps RMS of added jitter, and the pulse generator adds an additional jitter of up to 2 ps. The active amplifier may also contribute to the uncertainty. In terms of vertical uncertainty, the picoscope contributes less than 2.5 mV RMS; the pulse generator and amplifier also likely have a contribution but the values for these are unknown. All of these above-mentioned factors are random sources of noise.

As in [9, 12], the contribution of horizontal and vertical uncertainties to the measured received signals is described by the following equation:

$$X_i = x(t_i + \tau_i) + \eta_i \tag{1}$$

where X_i is the measured received voltage at sample i, $x(t)$ is what the signal would be without any noise present, t_i is the time at the ith sample instant, τ_i is the random jitter at sample i, and η_i is the random additive noise at sample i. The jitter τ_i and the noise η_i are both assumed to be independent zero-mean random variables. In our measurement scenario, the recorded τ_i is the net sum of the jitter due to all of the factors mentioned above, and η_i is the net noise due to all vertical uncertainty contributing effects.

The physical factors that could affect measurement repeatability are also numerous. In [13] we saw that antenna movement within the radome can cause up to 11 mV of additional vertical noise (on the same order of magnitude as the signal for some transmit-receive antenna pairs). Thus, for test scenarios presented here, we never move the antennas. However, there are other physical factors that cannot be

controlled. For instance, each time a patient comes for a scan the breast is inevitably positioned in the radome in a slightly different way than it was before. Thus, the distance between the antennas and the breast surface is variable, and so is the ultrasound gel thickness and its distribution around the breast. Further, if the ultrasound gel is replaced between patient visits this could lead to additional variations due to the precise volume of gel used, and a larger change in the distribution. Precisely because such factors are difficult to control, it is critical that we are aware of them so that we can minimize them in our next-generation system design. Measurement uncertainties due to the noted physical factors are not random: for any single measurement the outcome neglecting random noise is the same. However, in between measurements the physical factors vary introducing an uncertainty in the differential signals that is neither random nor systematic.

4. MEASUREMENT SCENARIOS

This section describes the measurement scenarios. Section 4.1 briefly summarizes the testing on phantoms reported in [12], 4.2 describes measurements performed with immersion medium, and finally 4.3 describes testing through clinical trials. The studies in 4.1 and 4.2 investigate measurement uncertainty and their effects through simplified testing scenarios. The tests in 4.3 examine the system use in a fully realistic scenario.

4.1. Testing with Phantoms

Prior to testing our clinical interface with patient volunteers, we studied the measurement uncertainty of our system with tissue-mimicking phantoms [12]. Breast phantoms offer known position relative to radome, known range of dielectric properties, known and consistent surface shape, no need for an immersion medium, no movement during or between scans, and scans can take as long as needed and be repeated as frequently as desired. The following two subsections, 4.2 and 4.3, will each add to the complexity of the testing setup towards a realistic scenario of testing with patients.

4.2. Testing with an Immersion Medium

We perform several scans with the system filled with the immersion medium (without a breast or breast model present). These scans allow us to gather information on the base noise levels that can be expected in our clinical system, without additional factors. Thus, we are able to characterize the amplitude and time uncertainties that are directly related to the system operation. The uncertainties, however, will include effects due to the ultrasound gel's conformity with the radome and location with respect to the antennas.

Measurements obtained using ultrasound gel cover several scenarios: same-day measurements, day-to-day measurements, and measurements that involve movement or replacement of the ultrasound gel. The first two scenarios will allow us to see the base noise levels in the system when it is left undisturbed, whereas the third scenario is one step closer to a realistic situation in which the ultrasound gel (and breast tissue) will appear differently in different scans due to varying ultrasound gel thickness, positioning of the breast relative to the ultrasound gel and to the radome walls.

For each measurement we take two successive scans, one with 16 hardware averages on the picoscope and one with 32 averages. The more averaging, the less the random vertical uncertainty in the received signals; however, averaging significantly adds to the scan time. The resulting data suggests minimum uncertainty levels that could be expected when patients are involved.

4.3. Testing with Healthy Volunteers

Finally, we test the system with patient volunteers. Performing breast scans on real patients adds to the complexity of the test scenario in multiple ways: the tissue properties and their distribution is unknown, the location of the breast surface (skin) is hard to determine precisely as it no longer coincides with the radome wall, and the use of the immersion medium implies that the distribution of ultrasound gel around the breast may vary when the breast is repositioned in the radome. Further, patient movement is possible during scans and positioning of the breast in a repeatable manner is challenging.

As is typical with early clinical trials, the patients we scan in this test scenario are, in fact, all healthy volunteers. Scans on volunteers are useful for testing the system consistency and repeatability, as will be demonstrated in the results section. The data obtained with healthy volunteers encompasses 342 breast scans taken over 57 patient visits. The breast scans are recorded for 26 different breasts (13 patient volunteers) over an eight-month period. Over the study duration, each volunteer was scanned up to once per month. The final data set includes scans from volunteers who visited a minimum of two and a maximum of six times. Volunteers have cup sizes from A to D, and are aged from 21 to 77. None have a personal history of breast surgeries or breast cancer.

Each set of breast scans contains three scans taken from the same breast of the same patient during a single visit. In particular, the first two scans are taken consecutively with no changes in between. This allows us to confirm that the base level of random noise (without the effects of breast movement or position) does not affect the resulting images. The third scan is then taken after the volunteer's breast is repositioned in the radome. This mimics the effect of a patient having scans on different days — the breast is inevitably in a new position in the radome, with (possibly) a new distribution of immersion medium around it. Since the goal of our system is breast health monitoring, it is important that the scans be repeatable over time regardless of these conditions. Thus, the third scan enables us to analyze how much the breast position and immersion medium affect the corresponding reconstructed breast images. In the following discourse, we refer to the first two scans as "Baseline1" (BL1) and "Baseline2" (BL2), respectively; whereas the third is labeled "AfterRepos" (AR) to indicate that the scan takes place after the breast is repositioned. Similarly, the left breast is denoted as "L" and the right breast as "R". We further note of the 342 breast scans recorded, half (all right breast scans) were recorded with the oscilloscope set to take 16 hardware averages and the other half (all left breast scans) with 32 averages.

An initial study comparing the after repositioning scans to baseline scans was presented in [14] that included data from four volunteers and compared images for one select volunteer. Here, we expand the study greatly, with a wider range of volunteers, and a more in-depth investigation of the data.

5. METHODOLOGY

The format is consistent for all recorded data. Each scan contains 240 signals, one from each transmit-receive antenna pair. The data is obtained by selecting a transmitting antenna, and cycling through with the remaining 15 antennas receiving; then the transmit antenna is switched and the process repeats for a total of 16 transmitting antennas with 15 receiving antennas each ($16 * 15 = 240$). Each signal is 1024 samples long and is collected at an equivalent-time sampling rate of 40 GSa/s. Breast scans may be recorded by taking either 16 or 32 averages of each received signal. Scans using 16 averages take 81 s to complete, and scans using 32 averages are 121 s long.

If the scan data is to be used for image generation, it undergoes several post-processing steps. The data is filtered, windowed, and time-aligned. In [12], we introduced two possible methods for time-aligning the data to mitigate jitter, a correlation alignment method and a method that aligns based on the reference (clock) signals. Based on the results in [12], we apply both in the study presented here.

The compensated signals are input into an image-generation algorithm. In particular, we use the Delay-Multiply-and-Sum (DMAS) algorithm [17], which is a relatively simple method commonly used for this application. The images are 2-D slices that can be stacked to obtain the 3-D image. They show regions of electromagnetic scattering within the breast. As with all imaging algorithms, the imaging parameters can have a serious impact on the resulting images. For this series of tests, we maintain all imaging parameters constant so that this source of uncertainty does not affect our study. Further, the accuracy in estimated breast tissue properties will in turn affect the accuracy in the reconstructed images. We calculate average breast tissue properties for each breast by identifying the time delay for different paths of propagation through the breast; this can then be used to estimate the corresponding tissue properties. The average breast tissue properties used in the imaging algorithm are the same for each scan of a given breast for each patient.

Two types of images are generated for ease of comparing the scans: "differential" images and "difference" images. Differential images are obtained by subtracting (after time-alignment) a calibration scan from the scan of interest. The differential signals are then input into the imaging algorithm. This

method allows for removal of artifacts that are always present in the system, regardless of if there is a breast present in the radome or not. For this study, the calibration scan is obtained by filling the radome completely with the ultrasound gel immersion medium. Difference images, on the other hand, are the result of subtracting two already-generated images. This allows us to visualize the similarities and differences between the two scans.

6. RESULTS

6.1. Measurements with an Immersion Medium

We first present results of system operation without a breast or phantom present (i.e., the radome has only the ultrasound gel immersion medium in it). In Figure 3 we plot data from a sample transmit-receive antenna pair that was recorded within a single day. Within this day, 10 consecutive scans were taken, five with 16 averages and five with 32 averages. These raw signals are seen to have some variation in amplitude and phase, but clearly have the same shape and are well aligned in time.

Next, we plot data over multiple days (again from a sample transmit-receive antenna pair). Figure 4 shows the raw, unprocessed signals collected over five days of measurements. As before, each day's measurements contain five scans with 32 averages and five scans with 16 averages. We further note that on three of the days, the ultrasound gel was removed from the radome and replaced (with same volume) before continuing with measurements. From Figure 4, it is seen that the variation in both amplitude and time shift are significantly more over multiple days than on a single day. We note that there are visible late-time discrepancies between the signals (between approximately sample 495 and 510) — these are attributed to the ultrasound gel replacements. Figure 5 shows the successfully time-aligned data of the same scenario. The signal alteration due to the three ultrasound gel replacement events is seen more clearly with changes in both amplitude and shape.

Further, it is found that when the equipment is first turned on, before reaching a steady state, the time uncertainty can be as high as 17 samples (425 ps) and the amplitude uncertainty as much as 61 mV. For data collected under these conditions, our time-alignment method frequently fails because the signal shapes vary significantly. Thus, it is of utmost important that the equipment is given sufficient warm up time before use; here, we determine that after two hours the uncertainty is minimized.

We next summarize the uncertainty seen in scans of the immersion medium. Table 1 shows the maximum uncertainty in amplitude and time for scans taken on a single day, over multiple days, and when the ultrasound gel immersion medium is replaced in the radome. We note that the amplitude uncertainty is calculated as the absolute value of the maximum differential signals, after the time misalignment is compensated for, such that only true amplitude variations are represented and not
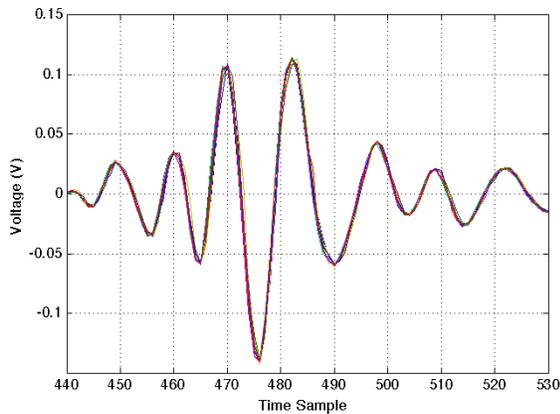
Figure 3. Plot of raw data from a select transmit-receive antenna pair for multiple scans of the immersion medium within one day.
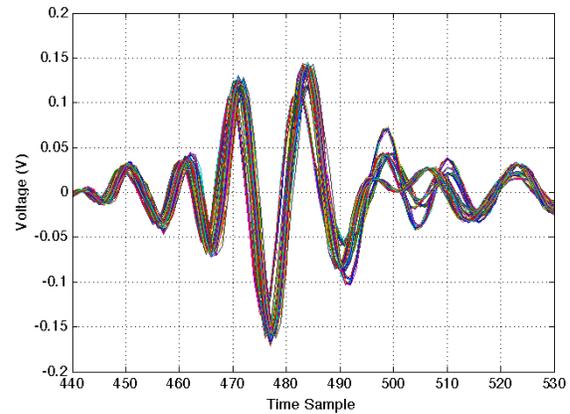
Figure 4. Plot of raw data from a select antenna transmit-receive pair collected from multiple scans of the immersion medium over a span of several days.

Table 1. Maximum uncertainty in signal amplitude and time for scans within a single day, over multiple days, and with ultrasound gel replacement. Results for the amplitude are shown for 32 and 16 averages; the time shift is shown before [after] time-alignment compensation is applied.

	maxΔV (mV)		maxΔt (samples)
	32 av	**16 av**	**[after alignment]**
same-day	2.4	2.9	1 [< 1]
multiple days	2.5	3.4	3 [< 1]
with replacement	43		2 [< 1]
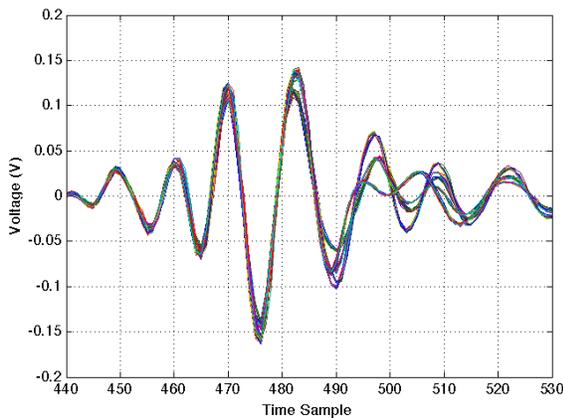
Figure 5. Plot of time-aligned data from a select antenna transmit-receive pair collected from multiple scans of the immersion medium over a span of several days.
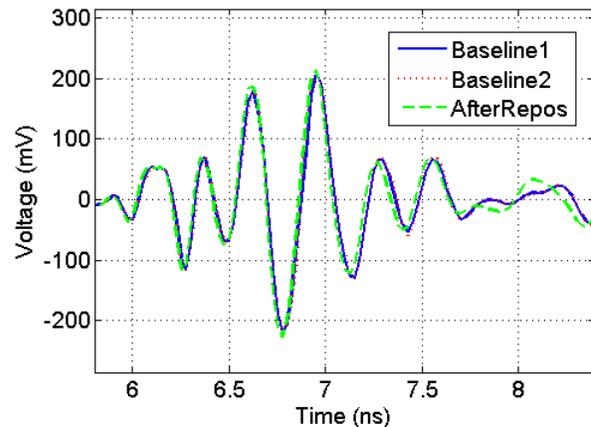
Figure 6. Plot of raw, unprocessed data for a select transmit-receive antenna pair, for three scans on the same patient: Baseline1 (blue, solid), Baseline2 (red, dotted), and AfterRepos (green, dashed). The Baseline scans are taken consecutively, while the AfterRepos scan is taken after the patient is repositioned.

variations due to timing shifts. The data shows that both horizontal and vertical uncertainty is small when data is recorded on a single day: 2–3 mV of vertical noise (depending on the number of averages), and at most 1 sample offset due to horizontal uncertainty. Over multiple days, the vertical noise is slightly higher, while the horizontal noise increases to 3 samples. The effect of replacing the immersion medium increases the vertical noise, indicating that, for meaningful scan comparison of the same patient, the ultrasound gel should remain in place for all measurements. We note that in all measurement scenarios, the horizontal compensation techniques applied together reduce the time uncertainty to less than one sample.

6.2. Scans with Healthy Volunteers

In this subsection, we first examine the collected signals, and then we investigate the reconstructed images. We note that the ultrasound gel volume is consistent between all scans of one patient visit, but not in between visits. Further, in accordance with the results from Section 6.1, the ultrasound gel is not replaced in between the three scans of one visit, as it was seen to be a source of added noise.

In Figure 6, we plot raw, unprocessed received signals for a sample transmit-receive antenna pair from the three scans taken during a single patient visit. The Baseline1, Baseline2, and AfterRepos scans are all found to match closely, at least visually.

Table 2. Maximum, mean, and minimum cross-correlation values (over all patient visits) for {*BL2, BL1*} and {*AR, BL1*} of each breast. The left breast scans were performed with 32 averages (32 av) and the right ones with 16 averages (16 av).

	Left (32 av)		Right (16 av)	
	{*BL2, BL1*}	{*AR, BL1*}	{*BL2, BL1*}	{*AR, BL1*}
max	0.999	0.999	0.999	0.999
mean	0.961	0.931	0.940	0.922
min	0.927	0.592	0.882	0.680

Table 2 compares the collected data from one scan to the same data from a different scan. In particular, for a single patient visit, the maximum, mean, and minimum cross-correlations are calculated between each the following scans: Baseline1 and Baseline2, {BL1, BL2}; and Baseline1 and AfterRepos, {BL1, AR}. The cross-correlation is normalized such that the autocorrelations of each signal at zero time lag are equal to one, and is calculated for the two signals at the point at which the signal overlap is maximal. The results are shown for both breasts. Within each set of scans, the correlation is taken for each pair of signals from the same transmit-receive antenna pair and averaged, for a resulting one value per comparison. This is repeated for each of the 57 patient visits.

From Table 2, we see that the maximum cross-correlation is greater than 0.99 for all three scan comparisons. The mean is also above 0.92 in all cases, however the minimum is seen to be as low as 0.59. The low minimum values for {AR, BL1} for both 16 and 32 averages are caused by signals recorded from antenna pairs that are the furthest apart, such that attenuation through the tissues is at a maximum and background clutter levels are high. Overall, these results, obtained from 82,080 signals, indicate consistency in the recorded signal shape. The data also indicates that, as expected, the noise in breast scans recorded with 16 averages is higher than in those recorded with 32 averages as seen in a minor deterioration of the mean cross-correlation for {BL2, BL1} with 16 averages when compared to the 32 average case.

We next examine images reconstructed from the scan data. In Figure 7, we show differential images for two sample patient visits: P1 is the first patient visit, and P7 is the 7th patient visit (out of 57 total visits). Specifically, we show results for the left breast for P1 and the right breast for P7. Three images are shown for each patient visit: the differential images for the Baseline1, Baseline2, and AfterRepos scans. In these images, red represents regions of high electromagnetic scattering whereas blue indicates areas of weak scattering. The images are plotted with the color scale set to show the normalized energy level of each pixel, in linear scale. The differential images are shown post-compensation for horizontal uncertainties. We see that the three differential scans for each patient are visually similar, and little variation is noticeable. This confirms that not only is the random noise between the Baseline1 and Baseline2 scans minimal but so is the remaining noise (after compensation) due to the repositioning of the patient's breast.

We next investigate subtler differences between the three scans of each patient visit. In Figures 8 and 9, difference images are shown that correspond to the images in Figure 7. For each patient visit, two difference images are obtained: a) the difference between the baseline images (BL2-BL1), and b) the difference between after repositioning and the first baseline (AR-BL1). The difference images are plotted normalized to the peak of the corresponding differential images. In each of Figures 8 and 9, we show first the difference images in linear scale; these images are mostly blue, indicate little to no changes in between the two scans that the difference image was derived from. Since a low-level of noise is seen in the differential images, it is hard to quantify how much it could affect the images. To better highlight the differences in the original scans, we also plot the images in log scale. In ideal circumstances, the difference images should have zero content — because signals from the two scans they are created from should be identical. However, in actuality, some noise is found. From the log-scale difference images, it is easily noted that the difference between baseline scans is significantly less than the difference between the after repositioning scan and the first baseline. From Figures 8(b) and 9(b), it is evident that the noise level between baseline scans is between -15 to -20 dB below the peak of the original images.
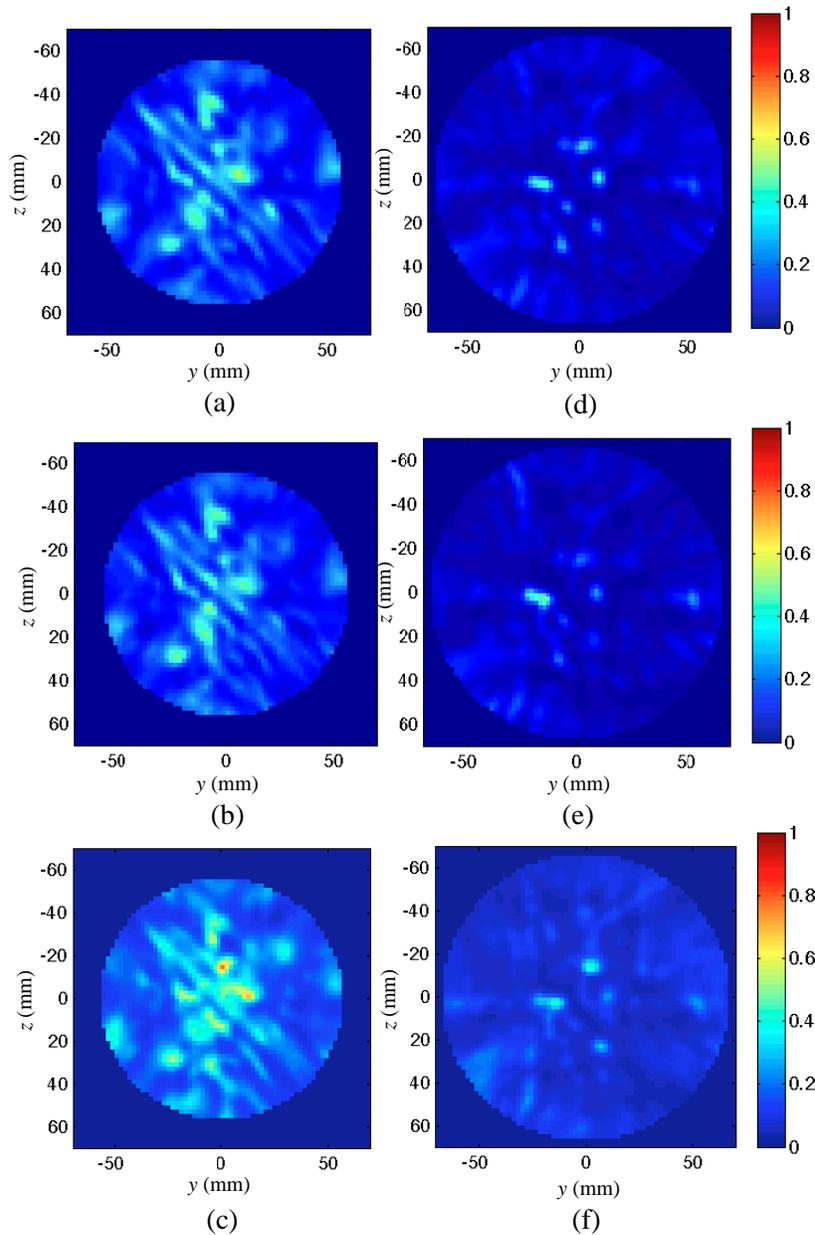
Figure 7. Differential images (a), (b), (c) for P1 left breast and (d), (e), (f) for P7 right breast; (a) and (d) Baseline1 calibrated with ultrasound gel; (b) and (e) Baseline2 calibrated with ultrasound gel; (c) and (f) AfterRepos calibrated with ultrasound gel. All images shown are after compensation for time-alignment of the data.

After the breast is repositioned, the noise level is seen to increase to -10 to -15 dB (from Figure 8(d) and Figure 9(d)). We note that modifying the parameters of the imaging algorithm can change the noise level seen in the images, which is why we keep the parameters consistent for all images; however, the noise introduced due to the algorithm is significantly below the level of noise seen here due to other factors.

Finally, in Figure 10, we provide a summary of the information shown in Figures 8 and 9 for all patient visits. We plot the mean difference between images, in decibels (dB), for all patient visits. The results for the difference between the second baseline scan relative to the first (BL2-BL1), and after repositioning relative to the first baseline scan (AR-BL1), are shown. From Figure 10, it is found that
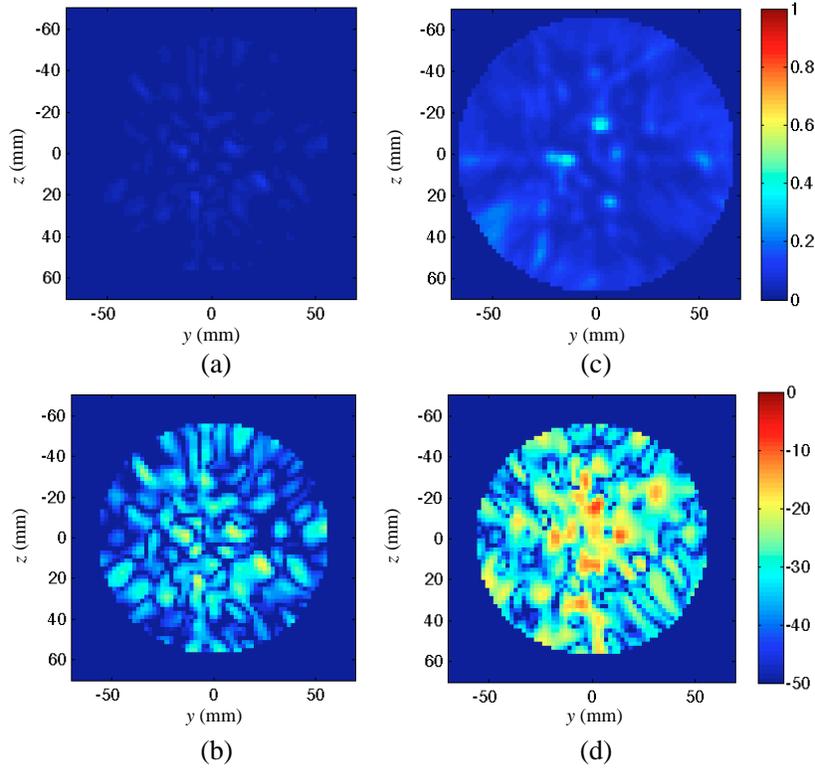
Figure 8. Difference images for the slices plotted in Figure 7 for P1 left breast: (a), (c) in linear scale and (b), (d) in log scale. Left images: BL2-BL1 (i.e., a difference between Figures 7(a) and (b)); right images: AR-BL1 (difference between Figures 7(a) and (c)).

the baseline scans are more similar to each other, and thus there is less noise in the difference images, than for the after repositioning scan compared to the baselines. However, the disparities are marginal: the median of the mean difference for all baselines is -36.4 dB while the median for after repositioning is -32.6 dB; and the worst-case highest mean noise level is -22.7 dB for baselines and -16.8 dB for after repositioning.

Lastly, we apply an image similarity metric to the three differential images generated for each breast for each patient visit. The similarity metric we choose to apply is called the Structural Similarity Index (SSIM) [18]. This metric is useful for comparing an image to a reference image in the same type of way that the human eye perceives image differences: by examining the relationships between the pixels that are spatially near to each other. This is opposed to other popular methods, such as mean-square error (MSE) or cross-correlation [19, 20], which compare intensities of pixels solely on a pixel-to-pixel basis. In particular, the SSIM takes into consideration three distinct image qualities: the luminance, the contrast, and the structure. First, the luminance is accounted for by calculating the mean intensity of each image and removing it to obtain zero-mean data. Next, the contrast is estimated as the standard deviation of the image values; each image is then normalized to their standard deviation to obtain data with unit standard deviation. These steps put both images into the same scale for contrast and luminance, allowing for a similarity comparison that is independent of these factors. Finally, the normalized images are compared structurally. The SSIM index value is equal to one if the image under test is identical to the reference image; it becomes zero when the two images have no similarity whatsoever.

Figure 11 plots the SSIM values for the left and right breasts for the following two images comparisons: 1) comparing the (differential) Baseline2 image to the reference image Baseline1 (red dots), and 2) comparing the after repositioning image to the reference Baseline1 (blue squares). The SSIM values are obtained by calculating the SSIM for each 2-D slice, and then averaging over all slices in order to obtain a single value per 3-D image. From the plots in Figure 11, it is clear that, in general,
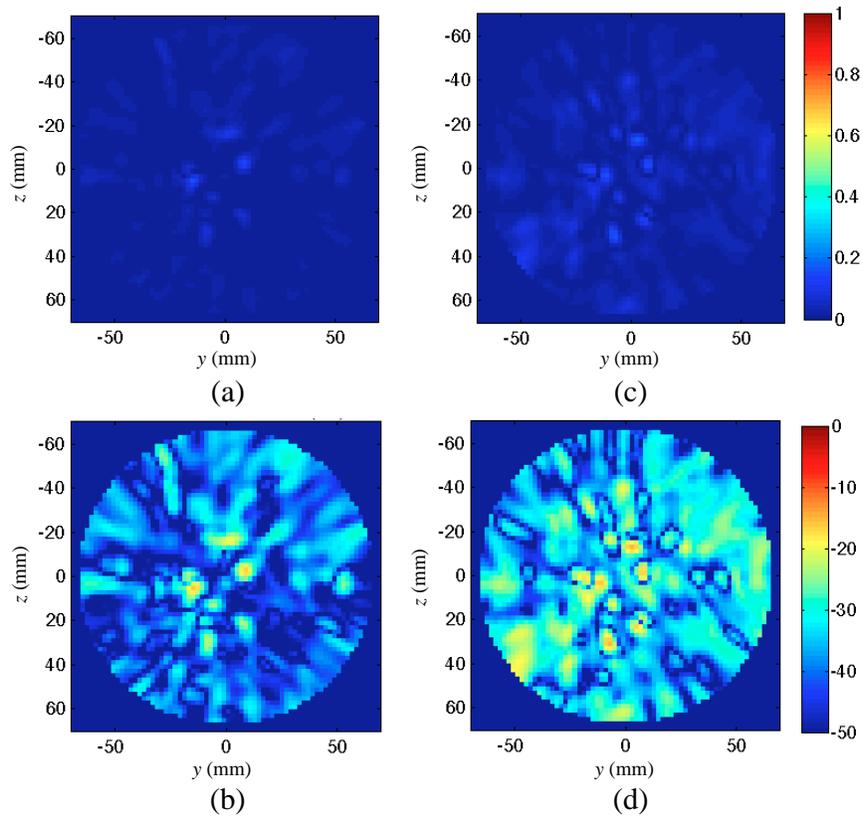
Figure 9. Difference images for the slices plotted in Figure 7 for P7 right breast: (a), (c) in linear scale and (b), (d) in log scale. Left images: BL2-BL1 (i.e., a difference between Figures 7(d) and (e)); right images: AR-BL1 (difference between Figures 7(d) and (f)).
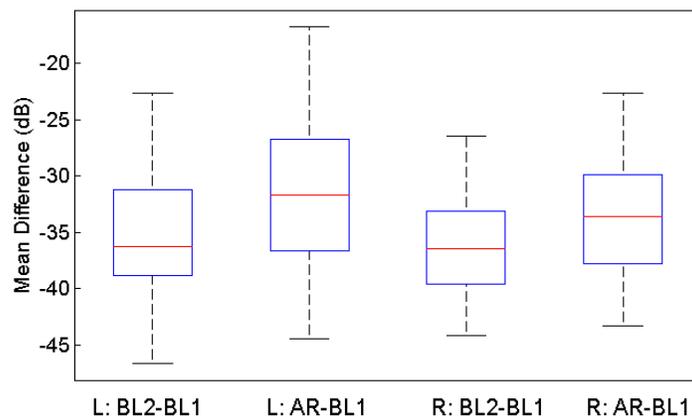
Figure 10. Boxplot of the mean difference in dB between Baseline2 and Baseline1 ("BL2-BL1") images and between AfterRepos and Baseline1 ("AR-BL1") images for each breast (L = left, R = right), over all patient visits. The red line indicates the median over all patient visits, the edges of the boxes are the 25th and 75th quartiles, and the whiskers denote the maximum and minimum calculated values for each scenario. In an ideal scenario (no noise), signals from two scans should be identical and thus the difference would be zero ($-\infty$ dB).
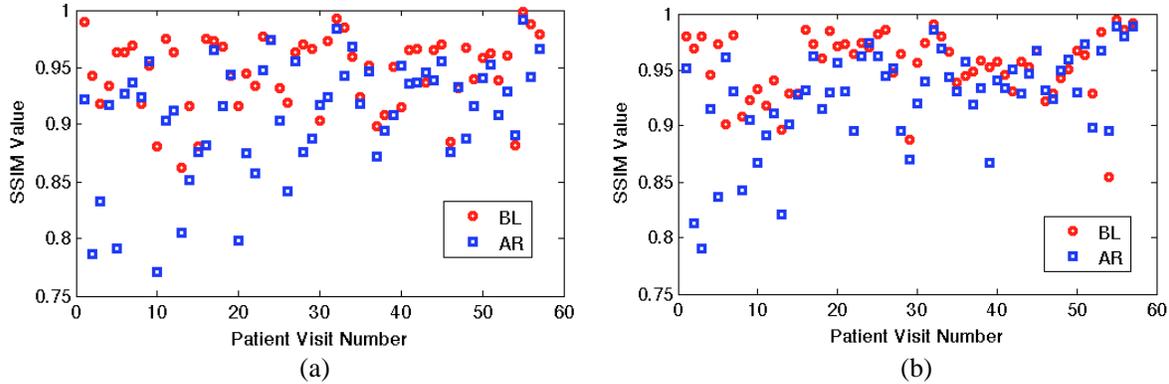
Figure 11. Plot of SSIM values for the Baseline2 images compared to Baseline1 (red dots) and for the AfterRepos image compared to the Baseline1 (blue squares) image for each patient visit; (a) left breast and (b) right breast. An SSIM value of 1 indicates that the images are identical, whereas a value of zero indicates no similarity.

the baseline images have a higher level of similarity than the after repositioning images does compared to the baseline scan. In fact, the average similarity (for both breasts) between baseline images is 0.95, while the average similarity for after repositioning relative to the Baseline1 is 0.92. While these are all high levels of similarity, it is unknown what amount of change the growth of a tumor would induce — this must be investigated with further studies.

7. DISCUSSION

The studies presented here confirm that the effect of horizontal noise outweighs the effects of vertical noise. In fact, as was seen in [12] with phantom testing, the difference between taking 32 averages and 16 averages for each signal in a breast scan is small. With either, the image is of good quality. However, further reducing the number of averages, or not having any averages, increases the background clutter seen in the images and may obstruct identification of a tumor if the breast has high glandular content. The series of tests confirms that breast scans can be run with 16 averages allowing for a scan time that takes almost half as much as the scan time for 32 averages (with immersion medium only, from Table 1: 3.4 mV vs. 2.5 mV of noise). We note that what is gained by the reduction in random vertical noise with 32 averages may be partially offset by patient movement during the required longer scan time.

From Table 1, the horizontal noise was found to be small when scans were taken over only one day (around 1 sample); however, for breast monitoring applications we need to be able to reliably compare data from multiple days. Our measurements showed that over multiple days the horizontal uncertainty was as much as 3 samples (Table 1); over long periods of time the uncertainty could further increase. We found that applying our time-alignment methods decreases the uncertainty between any scans to much less than one sample.

From the scans performed only on immersion medium it was found in Table 1 that removing and replacing the medium can lead to vertical noise as high as 43 mV. This value is on the same order of magnitude as some of the transmit-receive antenna pair signals, and thus is clearly not acceptable. In order to eliminate this source of noise, we suggest that the immersion medium not be replaced over scans of the same patient. Since this presents a physical challenge (if patients visit on days many months apart, we will inevitably need to replace the immersion gel), we will investigate soft solids that would not need to be replaced nor would vary in volume. We also note that the noise due to variability of immersion medium replacement (Table 1) is much larger than the noise attributed to breast repositioning (Figure 6, for example) due to the fact that the breast has a limited range of motion: the patient lies face down on the exam table, with their head on the pillow, so breast movement is only possible within a small range.

Despite all of the contributing noise sources, we see from Table 2 that the measurements can be quite repeatable, and from the images in Figure 7 that the breast repositioning does not seriously

degrade the image in terms of visual interpretation. Figure 7 also demonstrates the importance of horizontal uncertainty compensation, showing that with compensation the reconstructed images are consistent over various scans. From Figure 7 to Figure 10 it is seen that repositioning of the patient's breast has a small effect the differential (calibrated) images visually, but can have a noticeable effect numerically. The mean difference in the images between before and after repositioning is about 2 dB, but for some patient visits it is much higher. Clearly, it is best to minimize all noise sources where possible so that tumor identification can be successful. Therefore, in the future, we will study methods for breast positioning that can provide data that is as repeatable as possible. It is also key to study what level of change in signals or images could occur due to the growth of a malignancy — unfortunately this is a very challenging task to do in clinical trials, but before monitoring can become feasible one must have an idea of what amount of difference between scans could be attributed to healthy changes and what amount to abnormal changes in the breast tissues.

8. CONCLUSION

In this work we evaluated the repeatability of our time-domain microwave breast screening system from the point of view of breast health monitoring. To our knowledge, this is the first study that has examined the reliability of a microwave time-domain radar system for use with actual patient volunteers. In particular, we used data from a large set of patient visits during clinical trials to verify the system repeatability, compare the noise seen under various circumstances, including due to the immersion medium, and determine how it affected the reconstructed breast images. The clinical study involved 13 patients that had breast scans performed on 57 different occasions, for a total of 342 breast scans. Our time-alignment methods were shown to reduce the horizontal uncertainty to such a level as that the images are no longer negatively affected by it. We also thoroughly investigated that measurement repeatability in between scans due to patient positioning in the system. We found that $\sim$2 dB of difference in images from different scans can be attributed to patient position, and as such care must be taken to replicate positioning from scan to scan. In all, repeatability between patient scans was seen to be high, suggesting promise for a monitoring application.

Near-future work towards verifying system feasibility includes studies in long-term repeatability of patient scans and variability of scans due to normal monthly tissue changes. Studies on the expected level of change in signals and images due to cancerous growth in the tissue are also needed in order to be able to differentiate healthy, regular breast changes from irregular ones.

REFERENCES

1. American Cancer Society, "Cancer facts & figures," 2012.
2. Karellas, A. and S. Vedantham, "Breast cancer imaging: A perspective for the next decade," *Med. Phys.*, Vol. 35, No. 11, 4878–4897, Nov. 2008.
3. Fear, E., P. Meaney, and M. Stuchly, "Microwaves for breast cancer detection?" *IEEE Potentials*, 12–18, Feb/Mar. 2003.
4. Amineh, R. K., M. Ravan, A. Khalatpour, and N. K. Nikolova, "Three-dimensional near-field microwave holography using reflected and transmitted signals," *IEEE Trans. Antennas and Propag.*, Vol. 59, No. 12, 4777–4789, 2011.

5. Grzegorczyk, T., P. M. Meaney, P. A. Kaufman, R. M. diFlorio-Alexander, and K. D. Paulsen, "Fast 3-D tomographic microwave imaging for breast cancer detection," *IEEE Trans. Med. Imag.*, Vol. 31, No. 8, 1584–1592, Aug. 2012.

6. Bourqui, J., J. M. Sill, and E. C. Fear, "A prototype system for measuring microwave frequency reflections from the breast," *International Journal of Biomedical Imaging*, Vol. 2012, Article ID 851234, 1–12, 2012.

7. Jiang, H., C. Li, D. Pearlstone, and L. Fajardo, "Ultrasound-guided microwave imaging of breast cancer: Tissue phantom and pilot clinical experiments," *Med. Phys.*, Vol. 32, No. 8, 2528–2535, Aug. 2005.

8. Klemm, M., I. J. Craddock, J. A. Leendertz, A. Preece, D. R. Gibbins, M. Shere, and R. Benjamin, "Clinical trials of a UWB imaging radar for breast cancer," *2010 4th European Conference on Proc. Antennas and Propagation (EuCAP)*, 1–4, Barcelona, Spain, Apr. 12–16, 2010.

9. Zeng, X., A. Fhager, P. Linner, M. Persson, and H. Zirath, "Experimental investigation of the accuracy of an ultrawideband time-domain microwave-tomographic system," *IEEE Trans. Instrum. Meas.*, Vol. 60, No. 12, 3939–3949, Dec. 2011.

10. Sabouni, A. and S. Noghanian, "The robustness of HGA/FDTD in the presence of noise for microwave breast cancer," *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, 1–4, Charleston, SC, United States, Jun. 1–5, 2009.

11. Zeng, X., A. Fhager, and M. Persson, "Effects of noise on tomographic breast imaging," *General Assembly and Scientific Symposium (URSI)*, 1–4, Istanbul, Turkey, Aug. 13–20, 2011.

12. Porter, E., A. Santorelli, and M. Popović, "Measurement uncertainties in differential radar applied to breast imaging," *Proc. 2014 IEEE Sensors Applications Symposium*, 6–10, Queenstown, New Zealand, Feb. 18–20, 2014.

13. Porter, E., E. Kirshin, A. Santorelli, and M. Popović, "Microwave Breast Screening in the Time-Domain: Identification and Compensation of Measurement-Induced Uncertainties," *Progress In Electromagnetics Research B*, Vol. 55, 115–130, 2013.

14. Porter, E., A. Santorelli, and M. Popović, "Time-domain microwave radar for breast screening: Initial testing with volunteers," *Proc. 8th European Conference on Antennas and Propagation (EUCAP)*, The Hague, The Netherlands, Apr. 6–11, 2014.

15. Santorelli, A., M. Chudzik, E. Kirshin, E. Porter, A. Lujambio, I. Arnedo, M. Popovic, and J. D. Schwartz, "Experimental demonstration of pulse shaping for time-domain microwave breast imaging," *Progress In Electromagnetics Research*, Vol. 133, 309–329, 2013.

16. Kanj, H. and M. Popović, "A novel ultra-compact broadband antenna for microwave breast tumor detection," *Progress In Electromagnetics Research*, Vol. 86, 169–198, 2008.

17. Lim, H. B., N. T. T. Nhung, E. P. Li, and N. D. Thang, "Confocal microwave imaging for breast cancer detection: Delay-multiply-and-sum image reconstruction algorithm," *IEEE Trans. Biomed. Eng.*, Vol. 55, No. 6, 1697–1704, Jun. 2008.

18. Wang, Z., A. C. Bovik, H. R. Sheikhm, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, Vol. 13, No. 4, 600–612, Apr. 2004.

19. Penney, G., J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D–3-D medical image registration," *IEEE Trans. Med. Imag.*, Vol. 17, No. 4, 586–595, Aug. 1998.

20. Di Gesù, V. and V. Starovoitov, "Distance-based functions for image comparison," *Pattern Recognition Letters*, Vol. 20, 207–214, 1999.