Breast Ultrasound Technology and Performance Evaluation of Ultrasound Equipment: B-Mode

E. Sassaroli, *Member, IEEE,* A. Scorza, C. Crake, S.A. Sciuto, *Member, IEEE, M. Park*

***Abstract*— Ultrasound (US) has become increasingly important in imaging and image-guided interventional procedures. In order to ensure that the imaging equipment performs to the highest level achievable and thus provides reliable clinical results a number of Quality Control (QC) methods have been developed. Such QC is increasingly required by accrediting agencies and professional organizations, however, these requirements typically do not include detailed procedures for how the tests should be performed. In this paper a detailed overview of QC methods for general and breast ultrasound imaging using computer-based objective methods is described. The application of QC is then discussed within the context of a common clinical application (US guided needle biopsy) as well as for research applications, where QC may not be mandated and thus is rarely discussed. The implementation of these methods will help in finding early stage equipment faults and in optimizing image quality, which could lead to better detection and classification of suspicious findings in clinical applications, as well as improving the robustness of research studies.**

***Index Terms*— Image quality, B-Mode, needle guided biopsy, quality control.**

I. Introduction

An overview of quality control (QC) clinical protocols for ultrasound (US) imaging and biopsy based on computer-based methods is presented. For the purpose of illustration, we will provide examples of QC of ultrasound breast imaging systems. We will also discuss how these QC protocols can be extended to research US imaging systems. These protocols have been developed for B-mode imaging. The advanced imaging modes (e.g. harmonic imaging, spatial compounding, 3D/4D imaging) are not currently part of QC ultrasound imaging clinical programs. This is because the methodology required for their evaluation is still under development.

For several decades after its introduction in the early 1950s, breast ultrasound was primarily used to distinguish between cysts and solid masses. Recent technological advances in digital beamforming, transducer broad-band technology, and signal processing have made possible a better description of solid masses as benign and malignant based on their US features. This improvement in characterization of solid masses has brought to the development of a lexicon for reporting breast ultrasound findings that correlates with the one used in x-ray mammography: the Breast Imaging and Reporting Data System (BI-RADS) [1].

The probes (transducers) used to identify breast masses and to guide biopsies need to be high frequency, large broadband linear arrays, which are optimized for near field-imaging. A linear array has a rectangular field of view that maintains its width close up to the transducer face and therefore is particularly suitable when the region of interest extends right up to the surface as for the case of breast. Typical linear array transducers used in breast imaging today have between 192 and 288 elements along the long axis. When scanning the breast, a linear 12–5 MHz probe is commonly used. However, in small-breasted women or for evaluating a superficial lesion, a linear array 17–5 MHz probe may be used [2]. A higher frequency and large broadband allows the formation of a narrow beam and a short transmit pulse leading to good spatial resolution however at the expense of reduced penetration caused by increased attenuation of the ultrasound beam. Conventional linear array probes are electronically focused along the scan plane which is the xz plane parallel to the long axis x of the probe (Fig. 1). However, focusing along the elevation plane, which is the yz plane perpendicular to scan plane along the short axis y of the probe is usually achieved through an acoustic lens placed on the probe surface. The beam width in the elevation plane is commonly named slice thickness. Most of the probes used for breast imaging have a short axis acoustic lens which is focused at depths between *1.0* and *2.0 cm*. As a result, it is difficult to visualize details in tissue at a distance greater than 3 and 5 cm in depth by the 17–5 MHz and 12–5 MHz probe respectively.

QC is not only necessary for patient and operator safety but is also essential for maintaining the performance of the equipment to the highest-level achievable and it is required by various regulatory and accrediting agencies. Basic recommendations relevant to the testing of US imaging

This paragraph of the first footnote will contain the date on which you submitted your paper for review.

E. Sassaroli is with the Focused Ultrasound Laboratory, Brigham and Women's Hospital, Boston, MA, 02115, USA (e-mail: esassaro@bwh.harvard.edu).

A. Scorza is with the Engineering Department, University of Roma Tre, via della Vasca Navale 79, Roma, Italy (e-mail: andrea.scorza@uniroma3.it).

S.A. Sciuto is with the Engineering Department, University of Roma Tre, via della Vasca Navale 79, Roma, Italy(e-mail: salvatore.sciuto@uniroma3.it).

C. Crake is with the Focused Ultrasound Laboratory, Brigham and Women's Hospital, Boston, MA, 02115, USA (e-mail: crake@bwh.harvard.edu).

M. Park is with the Department of Radiology, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, 02115, USA (e-mail: miaepark@bwh.harvard.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

systems in a hospital setting have been published by a number of professional organizations. In USA, these are the AIUM (American Institute of Ultrasound in Medicine) [3-5], the AAPM (American Association of physicists in medicine) [6] and the ACR (American College of Radiology) [7]. In Europe, there are: the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) [8] and the Institute of Physics and Engineering (IPEM) [9]. In USA, quality assurance (QA) of personnel qualifications and QC of equipment permit hospitals to gain accreditation or re-accreditation of their practice by an accrediting agency (e.g. ACR, AIUM). Accreditation is required to obtain reimbursement for imaging examinations of Medicare and Medicaid patients and similar requirements are imposed by private insurance providers.

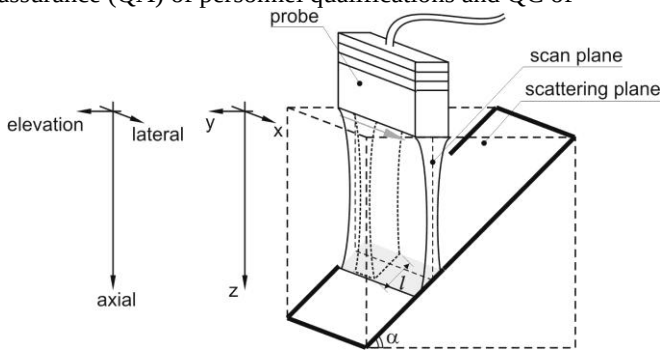

Fig. 1. Slice thickness by means of a scattering plane. The US beam is moved along the lateral direction to form the B-mode image from the scan lines.

The QC accreditation requirements by the ACR are listed in Table I. As of March 2016, some of these tests are required and some are recommended. These tests are the same for general and breast ultrasound, and also for US-guided breast biopsy. Required tests include: (i) annual survey of the *all* the ultrasound systems used in the practice and (ii) preventive maintenance by a service engineer. Recommended testing includes: (i) acceptance QC testing and (ii) routine QC testing. The acceptance testing of newly installed ultrasound equipment should at least include all the tests performed annually (Table I). The goal of acceptance testing is to confirm that equipment works according to the manufacturer specifications and it also provides a reference baseline to test that the equipment performance remains constant over time. The routine QC testing is recommended at least semiannually and includes all the required tests listed in Table I except for the system sensitivity test. Tests of uniformity, geometric accuracy, system sensitivity, contrast and spatial resolution must be performed using ultrasound test objects/phantoms. The ACR neither provides procedures on how to perform these tests nor specifies pass- fail criteria. Both subjective and objective methods can be used. If any test result falls outside the limit range established at acceptance by the medical physicist or manufacturer, corrective documented action by an equipment service engineer must be taken.

II. B-MODE IMAGING QUALITY CONTROL

The goal of a QC program is to detect a fault or a change in performance at an early stage so that appropriate technical help can be requested. The quality of ultrasound B-mode images depends on the probe, the associated electronics, the pre-and post-processing of the transmitted and received ultrasound signals and the display monitor fidelity performance. In addition, the so-called presets (ultrasound equipment settings) of the system which are chosen by the manufacturer and the preference settings used by individual users also influence image quality. Objective methods for B-mode are well documented in the literature (see for example [9-15] and references there) and are summarized in the guidelines recently provided by the EFSUMB [8] suggesting four levels of QC testing. Level 1: monthly, quick routine tests based on subjective methods with no test objects performed by sonographers. Level 2: semi-annual or annual extensive performance tests based on objective methods and the use of test objects. Level 3: acceptance testing also based on objective methods and test objects. Level 4: optional tests (or user requested) of probe temperature and acoustic safety indices: mechanical index (MI) and thermal index (TI) performed by the manufacturer, specialized labs or organizations (e.g.: FDA, USA; NPL, UK; TPB, GE). Level 4 is outside the scope of this paper and it will not be covered. Levels 1, 2, and 3 also require performance tests for the scanner and interpretation display monitors which are the same as for the other imaging modalities [16] and will not be discussed. All the QC tests should be documented.

TABLE I
ACR ACCREDITATION REQUIREMENTS

Required annual QC tests	Recommended annual QC tests
1. Physical and Mechanical Inspection	1. Geometric Accuracy
2. Image Uniformity and Artifact Survey	2. Contrast Resolution
3. System Sensitivity (maximum depth of visualization, SNR)	3. Spatial Resolution
4. Ultrasound scanner image display performance	4. Independent assessment of QC program (if applicable)
5. Primary interpretation display performance	

Ultrasound Equipment Settings

The ultrasound equipment settings for QC should be established at acceptance testing. Abdominal preset may be appropriate for general purpose US imaging systems and breast or thyroid preset for small parts ones. It is recommended to save the chosen preset as QC preset to avoid the risk that changes are made to the clinical preset. Settings such as: depth, frequency, dynamic range (compression), output power as given by the MI or TI, the overall gain, time gain compensation (TGC) should be established at this time for each performance test. The range of acceptable values for the performance parameters should also be established at this time. It is recommended (see for example [8-10]) to perform the QC tests by keeping the processing of the images to the minimum possible as it may mask problems with the equipment. B-mode post-processing features such as edge

enhancement (a spatial high-pass filter to enhance the appearance of anatomical features in the ultrasound image), persistence (a frame averaging to suppress random image noise) and reject (threshold value to reject the smallest echoes) should be set to the lowest level allowed.

The US image is an array (e.g. 1024×1024) of pixel grayscale echo values with usually 8 bits per pixel (256 values). These values are converted to grayscale luminance values by the display hardware. This conversion follows the DICOM Grayscale Standard Display Function (GSDF) which was created to standardize the display of image contrast and it is generally not adjustable [16]. Different grayscale maps, also called look-up tables (LUTs), provide different contrast in the image that all conform to the DICOM GSDF. For the measurements of some of the performance parameters, the use of a linear LUT is recommended. To verify that the preset selected for QC uses a liner LUT, the intensity of the grayscale wedge in the DICOM phantom image can be plotted as function of the number of pixels in the wedge (i=1,2, …n). If the LUT is not linear, a linear LUT has to be chosen for off-line analysis of the image. A method for generating a correction table so that the echo grayscale values are shown using a linear LUT is described in [10]. This linear LUT is stored by the software and used for the QC tests.

A. Level 1 – Performance testing by visual inspection

Level 1 is aimed at detecting by visual inspection: (i) any damage to the ultrasound equipment, especially to the probes and to their cables; and (ii) any significant change in equipment performance. In the literature, inspection of the mechanical integrity of ultrasound imaging systems has been reported to detect a significant number of scanner and probe problems. In the study by Hangiandreou *et al.* [17], 25.1% of faults were detected by visual inspection of US equipment over a four year period. Sipila *et al.* [18] reported a physical fault in 25% of the 135 probes and in 16% of the 51 scanners inspected. The tests performed at level 1 provide a relative measure of performance of the US system. A relative measure is one which tests the machine's performance against that of the same machine at another time, keeping the same US settings and scanning conditions (e.g. same light condition) [19]. These tests are obtained by acquiring in-air images without any phantom or coupling gel (Fig. 2a). The performance parameters evaluated at level 1 are: relative uniformity, relative sensitivity and relative noise. The relative uniformity test is primarily aimed at detecting faulty lens, cables, and a number of dead elements. The focus is moved close to the probe face, the overall gain is set to a high value to provide a uniform field with good signal at the probe face and the dynamic range set near minimum (high compression) for better contrast. The reverberation pattern should be uniform. Localized bright or dark regions may indicate defects in the transducer lens or a section of dead elements, for example (Fig. 2b). A number of consecutive dead elements can alter the power intensity pattern with a corresponding loss in ultrasound image [6] as also highlighted in the "coin test" [19]. This test consists in running a coin or an unfolded paperclip along the probe over a very thin film of coupling gel. The clip creates a vertical column of echoes which moves along the array as the paper clip moves. If the clip encounters a faulty element section, the column brightness will be reduced.

Sensitivity can be defined as the greatest depth at which speckle signals can be distinguished from electronic noise. Noise typically appears at the depth of the image as

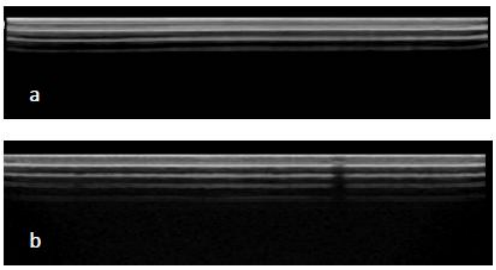

Fig. 2. (a) In-air image with a reverberation pattern. (b) Dark bands caused by probe surface damage or dead elements.

fluctuating 'low-level echoes' which do not persist between frames. An assessment of relative sensitivity can be made by determining the depth to which the reverberation pattern ends at the best setting of overall gain, field of view, frequency and focus in order to provide the deepest penetration into the ultrasound image.

The relative noise test can be performed with the same initial settings as for the sensitivity test [9]. Then, the TGC should be reduced to the minimum over the region of reverberation while keeping the TGC the same in the distal image where the noise is clearly visible. The overall gain should be reduced to the point the noise in the distal image disappears and this value should be recorded.

US imaging systems present often patterns of signal non-uniformity which are not equipment faults but are inherent to system design, image processing and minor imperfections. One way to recognize them is to see if the same probe model shows the same pattern.

B. Level 2&3 – Performance and acceptance testing by use of ultrasound phantoms and image analysis software

Levels 2 (performance testing) and 3 (acceptance testing) require phantoms and software-based performance tests. The ideal goal of these tests is to establish an absolute measure of performance which allows comparing performance of a machine with another one and against a standard specification. They include: accuracy in distance measurement, contrast and spatial resolution, uniformity and sensitivity. Acceptance testing should be performed for every US imaging system which enters the QC program. A list of QC software developed by companies and academic institutions is provided in [8] (e.g. UltralQ, QA4US).

1. Phantoms for QC of B-mode imaging

QC of B-mode imaging is usually performed with commercially available phantoms. A list is provided in [8]. Phantoms are made of materials which mimic the average acoustic properties of soft tissue (speed of sound, attenuation,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

backscattering). Embedded in these materials, there are targets which are designed to test US performance parameters. The background materials are usually either aqueous gels or rubber materials (e.g. urethane). Gel-based phantoms have a speed of sound at room temperature within 2% of the average sound speed of soft tissue (1540 m/s) and urethane rubber ones about 7% less [10] but they have the advantage to be more stable over time. The speckle pattern, similar to the one in soft tissue, is obtained by the introduction of small-diameter scattering material (e.g. graphite powder), which provides an acoustic attenuation in the range 0.5–0.7 dB cm^{-1} MHz^{-1}. This value depends on the amount of scattering material mixed with the background material. In addition, they have a backscattering factor in the range 1 to 4 $\times 10^{-4}$/m sr [10].

The effects of frequency and temperature on the acoustic properties of commercially available phantoms have been investigated in [20]. It was shown that the urethane rubber phantom has the highest non-linear attenuation response to frequency: $f^{1.83}$ and the greatest change in sound speed with temperature: 2.5 ms^{-1} $^0C^{-1}$. These facts need to be taken into account carefully when using the urethane phantoms especially for QC of high frequency transducers [21].

Commercially, there are multi-purpose phantoms with a number of targets for testing the different performance parameters and also phantoms for specific applications. For example, small parts or breast imaging phantoms are not as deep and have targets embedded at depths suitable for high frequency probes and often have wire targets with smaller diameters (0.05 mm). According to their function, targets may be divided in: point spread function (PSF) targets (or wire/filament targets), scattering plane target and grayscale (contrast) targets. The PSF targets are horizontal and vertical filaments of polymeric material (nylon) with typical diameter in the range between 0.1 mm and 0.3 mm (Fig.3). They are placed at various depths and are parallel to the short dimension (width) of the phantom scanning surface. They test the spatial impulse function and therefore the high contrast spatial resolution of the US system. The scattering plane target is a diffusive plane oriented at 45° to the top and bottom surfaces of the phantom [22,23] (Fig. 1). It is used to evaluate the high contrast spatial resolution in the elevational plane (slice thickness). The grayscale targets have known (nominal) contrast compared to the surrounding background material (e.g.: 9, 6, 3, -3, -6, -9 dB) with an accuracy usually not lower than ± 1dB. The background contrast is in the range between ±1dB. Typically, targets are cylinders of known diameters and location, with their length parallel to the width of the phantom (Fig. 6a). They allow the determination of the contrast resolution and local dynamic range. Phantoms made only of background and scattering materials are also available for testing the penetration capability and the uniformity of the speckle signal of the imaging system.

2. Accuracy in Distance Measurement

This is one of the acceptance tests suggested in the EFSUMB guidelines [8] and it should be also performed every time there is a software upgrade. No system problems were detected using this test by Hangiandreou *et al.* [17] in their 4 year study. The accuracy in distance measurement is usually evaluated by scanning the horizontal/vertical PSF filaments (Fig.3). This distance should be consistent with the clinical application of the probe used. When the scan plane is perpendicular to the filaments, they appear as spots in the image. The calipers are placed by the operator between two selected spots in the horizontal or vertical direction. Distance accuracy is assessed by comparing the measured distance between selected filament spots with their known distance.

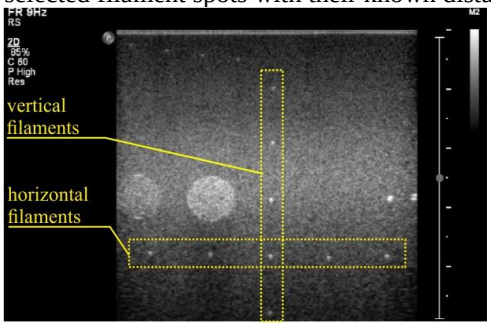

Fig. 3. Horizontal and vertical filaments in a multi-purpose phantom.

This is a subjective method and therefore is prone to errors especially for small distances measurements.

An objective method has been suggested in [14]. It consists in using off-line imaging software to measure the number of pixels between the centers of two selected filaments and to calculate the ratio of the measured number of pixels and nominal distance in mm. This measurement is repeated for several filament pairs, and ratios are calculated.

These ratios are then averaged to obtain the calibration factor in number of pixels/mm for the particular magnification, US systems and settings. This calibration factor can be used to calculate the distance in mm between any other two targets, knowing their pixel distance in the image provided that no changes are made to the image size, magnification and US settings. From these measurements, the relative error (in percent) can be also determined as $e_r=|d_k-d_m|/d_k\times 100$ with d_k and d_m known and measured distance respectively. Software for calculating the percentage error has been developed [24]. Given the image of the PSF filaments, the software asks the operator to choose n filament pairs by clicking on two filaments at the time which are separated either by a vertical or horizontal distance. It then calculates the absolute difference $\Delta_i=|d_{ki}-d_{mi}|$ for each filament pairs ($i = 1,2, \ldots,n$) as a function of the known distance d_{ki} and the linear least square fit method is employed to calculate the relative error e_r. This software has been validated using images of commercial PSF phantoms obtained from 14 US scanners, and 38 different probes. It was shown in [25] that the relative error e_r depends on sound speed, probe model, system settings (e.g. field of view (FOV)) and test method, with a range between 1.4 to 4.6%. The advantage of this method for evaluating accuracy of US imaging systems in distance measurement is its independence from the operator vision acuity and experience, and since it is automated, it saves time.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

3. Elevation Focus and Slice Thickness

The elevation focus cannot generally be adjusted by the operator, so it is important to establish its position. This is because when the scan plane transmission focus coincides with the elevation focus the sensitivity of the US imaging system is at its highest level. The depth of the elevation focus and the slice thickness for a given imaging system should be established at acceptance testing and do not to need to be repeated. A scattering plane phantom is required for this test if the method proposed in [22, 23] is used. The probe should be oriented so that the beam in the elevation plane (along the y direction) intersects the scattering plane at 45°(Fig.1). Echoes from the scattering plane are received from all image depths at which the transducer beam width intersects it. They are displayed as a horizontal band in the US system monitor, as illustrated in Fig.4. In particular, due to the 45° orientation of the scattering plane, the vertical (x direction in Fig. 1) and horizontal projections (y direction in Fig. 1) of the insonified portion of the scattering plane are equal to each other. The vertical height of the insonified portion of the scattering plane gives an *approximate measure* of the beamwidth at the average image depth of the scattering plane [22]. Actually the horizontal projection is equal to the beamwidth only if the beamwidth is changing slowly with depth: this is usually the case especially in the focal zone [23]. Therefore, for α=45° $L=l\times sin\alpha= l\times cos\alpha$ is *approximately* the beam width in the elevational plane, i.e. the slice thickness, where l and α are illustrated in Fig. 1 and L is given in Fig. 4. To determine the slice thickness at the elevational focus multiple images should be acquired close to the focus of the lens, as shown in Fig. 4.

A number of computer-based methods for computing the slice thickness as a function of depth have been developed [8,10,11,24]. The average width L of the slice thickness near the (lens) focus can be measured (e.g. Fig 4 time t_2). This average is obtained by selecting a ROI (region of interest) surrounding the slice thickness. The ROI selects a pixel matrix $M(i,j)$ with i pixel row index and j column index. The pixel values along each row are averaged and the plot of these values versus j index is the average width. A threshold is applied to the plot to determine the slice thickness. Usually, the threshold is chosen as the width of the plot corresponding to the full width of half maximum (FWHM) or the full-width-at-tenth-maximum (FWTM). The analysis is repeated for the other stored images having bands at different depths and a plot of the slice thickness as a function of depth is obtained. If the two slices, lower and upper, having a thickness 2L at FWHM are chosen, then the elevational focal zone can be evaluated as the distance δ between the centers of them (Fig. 4). Alternatively, one can find the minimum of the plot of the slice thickness as a function of depth. This minimum is the position of the elevation focus and the corresponding thickness is the slice thickness at the elevational focus. At the elevation focus, the slice thickness is around 1.5-2 mm for the 12-5 MHz breast imaging probe and 1-1.5 mm for the 17-5 MHz probe.

A method for estimating the elevation focus and slice thickness using the vertical filaments of a multi-purpose phantom was proposed by Skolnick [26]. The images of the filaments are acquired with the scan plane rotated at 45° angle to the filaments. The lateral width of the filament spots indicates the slice thickness at a given depth and the FWHM can be chosen as their width. For the measurement to be accurate the beam width in the elevation plane must be at least as great as the beam width in scan plane, which is usually the case. It might be possible for the beam width in the scan plane some distance away from the scan focus to be wider that the one in the elevation plane. In this situation, the beam width for the elevation plane represents the wider beam width of the scan plane. While the scattering plane phantom method offers its accuracy, the advantage of the Skolnick's method is that it does not require a special phantom.

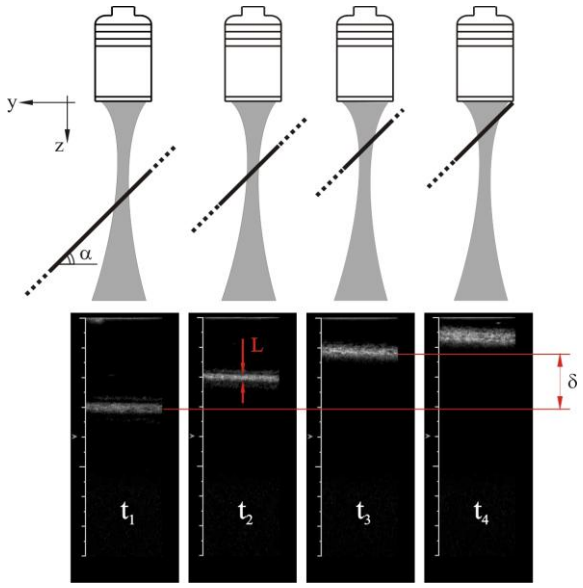

Fig. 4. Images of the scattering plane at different depths and focal zone.

4. Contrast Resolution and Local dynamic Range

The test for contrast resolution and local dynamic range should be performed at level 2 and more extensively at level 3 of QC. This test is aimed at establishing the ability of the US imaging system to detect subtle differences in the echogenicity of two targets. A linear LUT should be used for this test. The local or effective dynamic range can be thought as the range of dB values corresponding to the straight-line portion of the curve describing the relationship between the pixel grayscale values and the echo signal strength in dB. This curve is named grayscale mapping (GMF) function [3, 27] (Fig. 5). The slope of the straight portion of the GMF is the contrast or gamma of the imaging system.

The local dynamic range and the contrast vary with operator control settings, such as log-compression, pre-processing, transmit power, gain and the chosen post-processing LUT. To better understand this dependence, it may be helpful to review some ideas on B-mode image formation. The echo coming from a target located at a given depth is converted by each receiving piezoelectric element of the probe into a small voltage which is pre-amplified and digitized by each element analog to digital converter (ADC). At 12 bit resolution, this voltage can assume any of 2^{12} values so that an ideal dynamic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

range of 72 dB is available between the strongest and the weakest echo. The beam former sums in phase all the voltages after applying a suitable delay for each element. The resulting voltage is amplified by the TGC to correct for medium attenuation. The voltage is then rectified and smoothed (demodulated) by passing it through a low-pass filter, which removes the high frequency oscillations and retain the slowly varying envelope. Only the peak amplitude of the voltage envelope is used for image formation and that is the echo signal. The dynamic range for this echo signal is reduced through log-compression by the log-amplifier and digitized to 8 bits. The log-amplifier applies more gain to the weakest

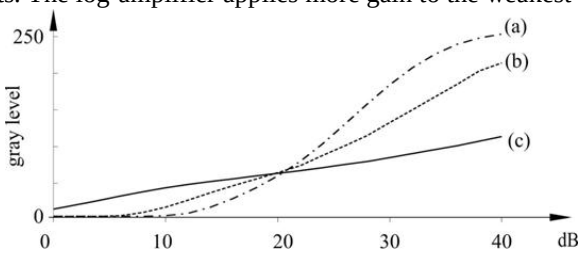

Fig. 5. Grayscale Mapping Function as a function of echo signal strength (dB). Effective Dynamic Range is between the two saturation zones of the GSM function. Compression settings: (a) 20 dB, (b) 40 dB, (c) 90 dB.

echo signals (coming from scattering within the tissue) than to strongest echo signals (coming from tissue interfaces) so that both echoes can be displayed at the same time. The log-compression can be adjusted by the operator by applying different types of compression, for example: low compression (80 dB dynamic range), medium compression (60 dB) and a high compression (40 dB). For instance, increasing the dynamic range setting from 40 dB to 80 dB enhances the signal strength from small echoes making the tissue echotexture or speckle pattern more visible but it also increases the noise. The compressed signal echoes are then assembled by the scan converter into a matrix with usually 8 bit resolution so that up to 256 grayscale pixel values are available to quantify the echo-signal. Before storing the echo signals in the image matrix some pre-processing is applied to improve image quality, such as: edge enhancement, persistence, zoom, compound imaging, etc. After image storage, the pixel values are interpolated to fill the empty spaces in the matrix. The display hardware provides the digital LUT that converts the pixel values to luminance grayscale values [16]. The user can choose among a number of LUT tables according to the contrast desired. A linear LUT preserves compression and the grayscale luminance values represent the pixel grayscale echo values. Compression and digitization to 8 bits of the echo signal is a non-linear process which is represented by the GSM function in Fig. 6 for different compressions. The GMF has a toe, a linear portion and a shoulder. The toe and the shoulder are saturation regions: weak or large echoes produces very little change in the grayscale values and therefore represent areas of low contrast in the image. In the linear portion, a small change in the echo amplitudes induces a visible change in the grayscale values, which corresponds to the diagnostically useful values as they produce the largest contrast.

Grayscale targets (Fig. 6a) are used to determine the contrast and local dynamic range of an US ultrasound system for a given compression. It is advised to use a phantom with at least 5 grayscale contrast targets (cylinders) of nominal contrast from -10 dB to 10 dB in steps of at least 3 dB with respect to background material. The measurements proceed as follows. For level 3, each cylinder is scanned five times at slightly different positions of the probe with the cylinder centered in the image as close as possible to the elevation focus depth. For level 2, only one acquisition is required for each cylinder. A circle is interactively fitted into the displayed disc. To avoid effects of shadowing caused by possibly greater

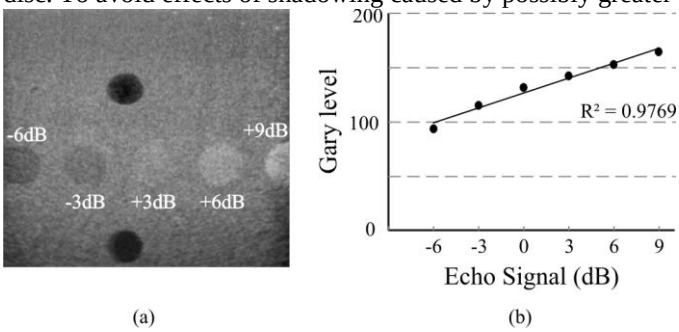

Fig. 6. (a) Grayscale contrast targets in a general phantom. (b) Grayscale pixel values versus nominal target contrast values.

attenuation within the circle, it is recommended that the mean of the gray levels within the top semicircle is estimated for each cylinder [3, 10, 15]. The ensemble mean and standard deviation of the four most consistent measurements are plotted for each cylinder as a function of the nominal contrast value in dB and a linear regression is made (Fig. 6b). The slope of this line, i.e. the grayscale echo values per dB, yields the contrast of the system at the given settings used for the measurements. It also provides the local dynamic range. This represents the linear part of the GMF. The grayscale phantoms are embedded with a few contrast targets, therefore only a part of the GMF can be plotted without extrapolation and it is not possible to evaluate directly the local dynamic range of the US imaging system that often ranges between 30dB and 100dB.

In [27], a novel method is proposed to determine the GSM function. It consists in acquiring a number of images of the same grayscale target at different gains in dB. The data are then interpolated according to a sigmoid function and the local dynamic range is determined. In comparison to the above described method, which evaluates the GMF with a few points, the proposed method allows evaluating the GMF with a higher number of points. Although the preliminary results are encouraging; other studies are currently underway as the developed procedure as the developed procedure can provide estimation of important parameters of system performance such as contrast resolution, effective dynamic range and sensitivity.

5. High Contrast Spatial Resolution

High contrast spatial resolution (HCSR) is defined in [3] as the minimum distance between two identical point targets which produces, for a given gain setting, a higher level of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

back-scattering than their surrounding medium permitting their individual identification. The point targets must be aligned with the direction of US propagation (axial resolution) or normal to it (lateral resolution). This performance parameter should be tested at levels 2 and 3 of the QC program. The lateral resolution at the focal depth depends on the wavelength λ of the ultrasound beam, the depth z and the width D of the active probe elements as $w= \lambda z/D$ [28].The diffraction limited axial resolution in the focal zone of the scan plane can be estimated as $d= \lambda M/2$ [28] where M is the number of cycles in the pulse. For broadband probes, the pulse is 1–3 cycles long giving an ideal resolution comparable to λ. A widespread subjective method to evaluate HCSR in the scan plane is to determine the axial and lateral resolutions on a series of nylon filaments which are separated by increasing smaller distances between two adjacent filaments in both directions lateral and axial (Fig. 7a). The spatial resolution is expressed as the smallest distance between any two filaments that can be differentiated in the vertical direction (axial resolution) and horizontal direction (lateral resolution). This method is subjective and affected by placement errors of the targets within the phantom, which can be significant for the smallest distances (higher resolution).

Alternatively, HCSR can be evaluated from the PSF of the imaging system in the scan plane (in-plane PSF) using an objective method: PSF is obtained by imaging the PSF filaments [4]. HCSR is affected by the position of the target in the image and by the dynamic range. It is recommended in [8, 10] to determine the HCSR of the imaging system at the elevational focus using a linear LUT. The vertical/horizontal filaments (Fig. 3) of the phantom are imaged with the elevation focus and the scan plane focus coinciding. The acquisition of the images of the filaments is repeated a number of times for estimating the precision of the in-plane PSF measurement and in these images, the filament closest to the elevation/scan focus is selected by placing a ROI around it (Fig.7a). An objective method consists in determining the FWHM or FWTM along the axial and lateral direction of the filament [10, 11, 25]. An example is shown in Fig. 7b where the grayscale values (levels) are plotted as a function of horizontal (lateral) distance. From the plot, the FWHM is determined. The FWHM along the lateral directions is determined for all the acquired images and the averaged and standard deviation is computed. This average represents the lateral HCSR of the US equipment system which is under evaluation. The same procedure can be applied to obtain the HCSR in the axial direction. The above measurements give the FWHM determined from the grayscale values. From the knowledge of the pixel values/dB, i.e. the contrast of the imaging system, the pixel values can be transformed in dB values and a plot of the dB amplitude values as a function of lateral or axial distance can be made. When the maximum of the plot is at 0 dB, the FWHM (-6dB) or the FWTM (-20 dB)

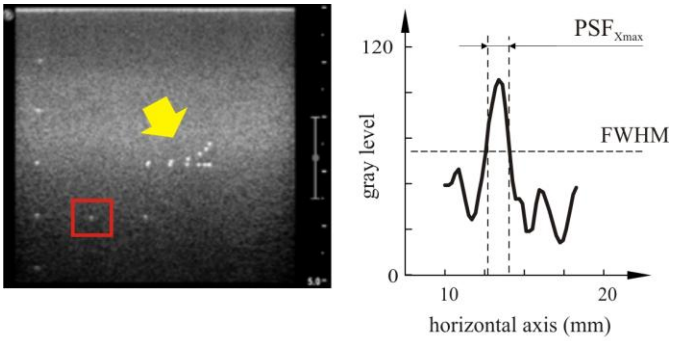

Fig.7. (a) Axial and lateral resolution group. (b) Lateral resolution at FWHM evaluated by means of the PSF filament within the ROI (red line) in (a).

width can be then obtained. For the high-frequency breast probes, at the focal zone (1-2 cm depth), the FWTM lateral resolution is in the range 0.5-0.9 mm and the FWTM axial resolution is in the range 0.4 to 0.7 mm.

6. Sensitivity

The maximum depth of signal visualization or maximum depth of penetration (DOP) quantifies the sensitivity of the US imaging system. The DOP can be defined as the greatest depth at which speckle signals can be distinguished from the electronic noise [3,4,15]. Speckle is a structured interference pattern generated by multiple small scatters. The depth of visualization of speckle is limited by the sensitivity of the probe/scanner combination. A computer based method based on the measurement of the image pixel SNR versus depth is described in the IEC report [15]. This method is recommended in [8, 30]. It consists in acquiring a set of image pairs using the same US settings. The pair includes one image of a uniform phantom (or a uniform region of a general phantom) and one in-air image acquired by decoupling the probe from the phantom. For the settings, the QC preset for the probe can be used with the depth and scan focus adjusted to optimize the visualization of the speckle signal disappearance and the output power and overall gain set to high values [15]. High compression (40-50 dB) may be also useful to better detect speckle signal and noise. For the image pair, a rectangular ROI is chosen on the central region of the image, spanning the near field to the bottom of the image (Figs. 8a and 8b).

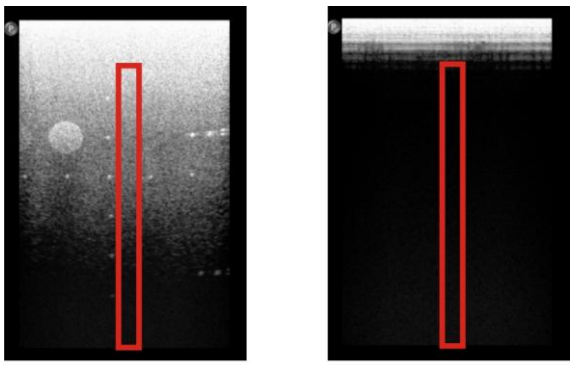

Fig. 8. Maximum depth of penetration. Images of (a) Uniform phantom and (b) in-air.

For both ROI, the laterally averaged grayscale pixel values as a function of depth is computed by the imaging software. The result is a pair of pixel vector: *A(j)* (phantom image) and *A'(j)* (decoupled probe image) with *j=1, 2 ,…., N* and number of rows in the pixel matrix inside the ROI. This computation should be repeated for 3 or more independent image pairs obtained by re-positing the probe at slightly different positions. A typical plot of these averages $\bar{A}(j)$ is shown in Fig. 9. As may be seen in the figure, the $\bar{A}(j)$ and $\bar{A}'(j)$ values gradually merge as depth increases. Assuming that speckle signal *S(j)* and noise are not correlated, then

$$\bar{A}(j) = \sqrt{S^2(j) + \bar{A}'^2(j)} \quad (1)$$

and the SNR as function is given by:

$$SNR(j) = \sqrt{\frac{\bar{A}^2(j)}{\bar{A}'^2(j)} - 1} \quad (2)$$

The value of *j* at which the *SNR =1* can be taken as the maximum DOP. Form the knowledge of the number of pixels/mm, the maximum DOP can be expressed in mm.

Hangiandreou *et al.* [17] found the sensitivity test not to be effective at detecting system problems as it detected only 1.6% of equipment failure. A novel method to calculate the maximum DOP is described in [29]. A ROI is selected in the penetration phantom and the laterally averaged grayscale pixel values are plotted as function of depth. The data are then interpolated with a polynomial function and the gradient of the fitting function is calculated as a function of depth. The depth at which the gradient is equal to a given threshold gradient, discussed in the paper, is the maximum DOP.

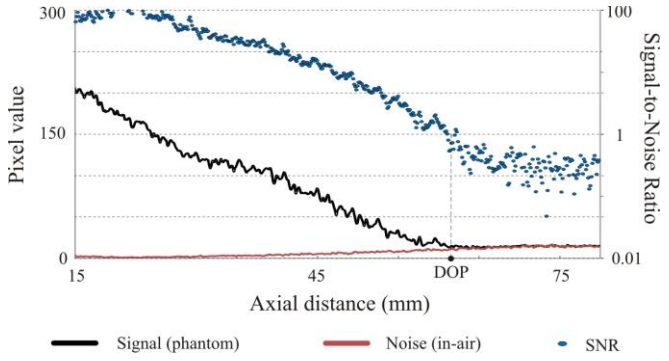

Fig. 9. Laterally averaged gray scale pixel value in (a) Uniform phantom, (b) In-air, and (c) SNR in axial direction.

7. Image uniformity

The uniformity test has been reported to be the most effective at detecting US equipment problems. In the same study mentioned earlier, Hangiandreou *et al.* [17] reported a 66.3% of faults detected by image uniformity inspection of more than 45 scanners and more than 265 probes over a four year period. Their method consisted in inspecting visually, images of a uniform region of a phantom and in-air images. Martensson *et al.* [31] using a commercial probe tester inspected 676 probes and found 39.8% of them defective. A probe tester measures the relative sensitivity of each element. The sensitivity test should be performed at levels 2 and 3 of the QC program using objective methods.

Faults usually appear as vertical dark line/band(s) (axial hypoechoic artifacts) indicating a localized transmission or reception fault. However hyperechoic artifacts have also been reported [32].

A computer based method is suggested in [8] and is based on the evaluation of the uniformity of both speckle and noise as described in the IEC report [13]. The uniformity of the signal (speckle) is checked by imaging a uniform phantom or a uniform region of a multi- purpose phantom. The uniformity of the noise is tested by acquiring in air-images. For these tests, the scan parameters such as depth, focus, frequency, transmit power, gain, TGC should be optimized to obtain a great sensitivity closer to the transducer face with dynamic

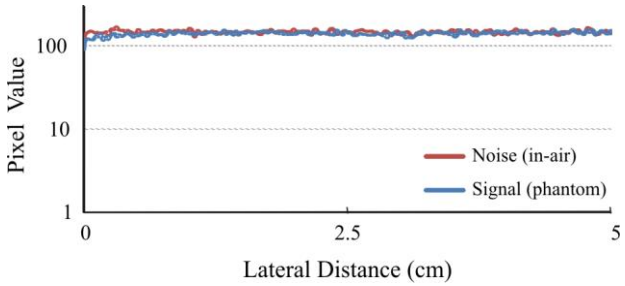

Fig. 10. Echo values versus width for noise and signal uniformity tests. The signal should be constant on average along the width except at the edges of the image where fewer elements are active.

range sets to near minimum for better contrast (40-50 dB). The procedure consists in acquiring an image of a uniform phantom and an in-air image. A superficial ROI is chosen in both uniform phantom image and in the in-air image. Given the pixel matrix *M(i,j)* inside the ROI, the pixel values for a given *j* (vertical distance) are averaged and the averaged signal pixel values are plotted as a function of *i* or lateral distance as shown in Fig. 10. If the gamma of the system is known, the signal values in dB can be plotted versus lateral distance. The noise should be constant on average: a decrease of a few dB (e.g. 1– 3 dB) within a few *mm* width or larger may indicate element and/or probe surface damage. The ROI dimensions should be kept the same and the signal and noise profiles should be compared from year to year.

Other methods have been recently proposed with the aim to quantify non-uniformities objectively [33,34]. In [33], the image gray level histogram of the ROI is weighted by sigmoid functions to highlight the spread of the image histogram due to non-uniformities. In [34], a pattern recognition algorithm is developed that divides iteratively the ROI into *m* non-overlapping regions (sub-ROIs) according to some textures f_k (*k=1, 2, 3, ...*) and are evaluated by means of co-occurrence matrices. The number of sub-ROIs and the dispersion of the f_k values are combined to measure the ROI non-uniformity. Although both methods have shown promising results, further studies are ongoing to improve both algorithms which also require a suitable choice of US settings.

III. Ultrasound Guided Breast Needle Procedures and Their QC

Efficacy of ultrasound-guided procedures depends critically on image quality of (a) the target organ and (b) the needle, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

on (c) the needle placement accuracy in real-time. Image quality of both the needle and the target organ can be related to ultrasound 2D quality assessment as described in Sec. II, nevertheless some specific features should be considered. In particular, requirements for the ideal needle for ultrasound guided procedure would include: (1) perfect needle visibility, in particular its tip; (2) suitability for all kinds of tissue; (3) perfect visualization at all angles; (4) sharp depiction of the rim of the needle; (5) no artifact formation; (6) no shadowing; and (7) extremely good detection and differentiation from the surrounding area. Although such an ideal needle does not exist, from considerations above, the image quality assessment in US breast imaging and guided biopsy could be performed evaluating needle visibility and placement accuracy through specific B-mode features, therefore in the following, an overview of main QC procedures for ultrasound imaging and biopsy based on objective methods and measurements is provided. A number of factors influence assessment of needle visualization [35, 36]: its size, orientation, mechanical treatments of inner and outer surfaces, motion, placement within the focal zones of transducers, shadowing, injection of fluids, presence of air or micro bubbles in solid tissues, US-scanner settings, use of new technological applications as such as coded excitation [36]. In the literature, some studies have been reported investigating reliability of US guided systems and methods, nevertheless most of them provide results that are operator dependent since they are based on a subjective assessment of needle visibility [37, 38]. In [38], six criteria were defined to describe the quality of different aspects of needle visibility in ultrasound-guided regional anesthesia: (1) needle visibility; (2) visibility of the surrounding area; (3) consistency of the needle surface; (4) formation of artifacts; (5) shadowing; and (6) the detection and distinction of the needle from the surrounding area. Criteria were rated using a categorical visibility score on a scale ranging from 0 to 10. In summary, every needle was described by 2 operators, in 2 media, at 2 angles, using 3 US machines, at 3 aspects of the needle and assessed by 6 criteria for every scan. Good needle depiction is usually expected in hypoechogenic tissue (similar to water) or when the ultrasound plane is perpendicular to the needle. Results show that needle visibility is reduced at steeper insertion angles [36]. Unfortunately, not every target can be reached with optimal ultrasonic depiction of the needle and there is a general complaint about decreased needle visibility with ultrasound at small (steep) angles [35, 39, 40], since in clinical use, angles between 30 degrees and 60 degrees are required. There have been many studies on echogenic needle visibility based on objective methods [41-44] by means of ultrasound phantoms (e.g. hydrogel phantoms). In [41], for the needle guide setting at a shallow insertion angle against the phantom surface, the needles were inserted 10 times to a depth of 35 mm in the phantom with bevels up, each time avoiding prior insertion sites. The captured images were stored digitally: objective visibility of the needle was digitally estimated as the difference in the mean luminosity between the needle area and the adjacent background (1 mm width) in the captured image. The mean luminosity was verified using a commercial computer software and was defined as a gray scale value between 0 (darkest, black) and 255 (brightest, white). The median contrasts of the 10 insertions for different insertion angles were calculated to determine the objective visibility. The region 1mm from the needlepoint was considered the tip, and the region 2-5 mm from the needlepoint was considered the shaft. Objective visibility was evaluated by separating the needle's tip and shaft. Two expert anesthesiologists judged the sampling positions by enlarging the image (i.e. 5x magnification). None of the echogenic needle tips were treated to heighten echogenicity. The regions 1-2 mm from the needle points were not evaluated because these regions include non-echogenic portions. The real-time assessment of needle position is very important, since a low accuracy in needle position may lead to a damage of collateral structures during the exam. (e.g. artery, vein, nerve). The needle placement should not be moved until every effort has been made to clearly display all these structures: this is a specific feature in US needle guided procedures, that require a clearly image display before and during needle insertion (the needle usually introduces artifacts in the image that are not present before) [35]. Although many articles in the scientific literature address the problem of the image quality and placement of the needle in US biopsy, the attempts to an objective assessment based on the ultrasound image quality features described in section II seem to be not widespread. To this aim in table II, we propose some tests that could be implemented: They might be a starting point for the development of methods suitable for the routine assessment of image quality in ultrasound breast imaging and needle biopsies. From table II, some features should be added to those described in section II: frame rate, duplex imaging features and image registration, power Doppler sensitivity. The frame rate is a very important parameter, since it is related to the capability to perform real time examinations and today it is one of the main advantages of US systems over other diagnostic modalities (e.g. MRI, X-Ray): US needle guided procedures must be performed in real time, therefore frame rate should be tested in the US system used. The above feature becomes critical if duplex imaging is used to display the perfusion of fluids during injections or blood flow and vascularization. A good frame rate (optimal also for cardiac imaging) is beyond 20 Hz, since below this value smoothing artefacts could become severe [45], nevertheless some studies suggest higher frame rates could be achieved by means of new beamforming techniques [46,47].

TABLE II – IMAGE QUALITY ASSESSMENT FOR ULTRASOUND BREAST IMAGING AND NEEDLE GUIDED BIOPSY

Critical issues	Image Quality test	References
Accuracy in needle placement within the tissues	a. Accuracy in distance measurements	[4-6, 9, 14, 24]
	b. High contrast spatial resolution (Axial and lateral resolution, Slice thickness)	[4-6, 9, 11, 12, 14, 22, 23, 26]
	c. Low contrast spatial resolution (Local Dynamic Range and Gray Scale Mapping Function, Contrast-	[4-5, 11, 15, 27]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

	detail curve, Lesion detectability) d. Maximum depth of signal visualization	[4-6, 9, 15, 29-31]
Uniformity of needle visualization	e. Image uniformity b. High contrast spatial resolution c. Low contrast spatial resolution d. Maximum depth of signal visualization f. Dead zone	[8, 9, 14, 18, 33, 34] [4-6, 9, 11,12,26] [4-6, 9, 15, 29-31] [4-6]
Real Time	g. frame rate	
Perfusion of fluids during injection	h. Color and Power Doppler Sensitivity l. Duplex imaging features and image registration (Color Flow B-Mode Image Congruency) g. frame rate	[53-54]

Moreover a lack of consistency between the image of the vessels and the corresponding Doppler signal can be significant for the accuracy in needle insertion during the guided procedure and biopsy [37] and, more in general, for a correct diagnosis (e.g. association of a vascular area to a lesion). However, although image registration is a very widespread topic in medical imaging [48], testing Color Image Congruency in Doppler ultrasound is rare and often qualitative [49-51]. A definition of Doppler sensitivity is reported in [52,53], as a measurement of the ability of an ultrasonic Doppler system to detect, above the noise level, a Doppler signal from a simulated point target (less than three wavelengths wide) of known target plane-wave reflection loss, moving at a specified velocity and placed at a specified distance from the probe. It depends mainly on nominal frequency, transducer type, clutter filter and output power setting, therefore it can be evaluated from penetration depth, lowest detectable velocity and vessel size, as it is described in [54] where all of them have been combined in a single figure of merit named sensitivity performance index *Si* equal to $Si=P_{LV}\sqrt{(D\times LV)}$ where D is the vessel diameter, LV is the lowest detectable velocity (i.e. the lowest velocity that can be displayed unambiguously), P_{LV} is the penetration depth of the lowest detectable velocity, which is the maximum depth in tissue from which a Doppler signal free from extraneous noise can be obtained. It was found that the limiting factor for the detection of the lowest velocity was found to be the vessel inner diameter, since a decrease in the diameter size resulted in a decrease in the color Doppler signal strength: this is likely why color and power Doppler effectiveness in cancer detection is controversial [55-57]. On the other hand we have not been able to find in literature an in-depth and comprehensive study that clearly correlate the performance of Doppler instrumentation (i.e. in terms of Doppler sensitivity, spatial resolution, color congruency, etc.) with the diagnostic efficacy in cancer detection. Before a definitive judgment, in our opinion more studies should be done to establish measurable quantities for performance evaluation of these instruments, specifying their settings and limitations and assuring objectiveness and traceability of results.

TABLE III
ULTRASOUND RESEARCH SYSTEMS, BASED ON [60] WITH ADDITIONS

Name	Type[a]	Key features	Refs.
Antares	C	Basic research functions. Allows access to RF data (after BF).	[61]
DiPhas	R	Scalable, modular system with SW beamforming.	[62]
Femmina	I	Modular back-end system to interface external BF to PC.	[63]
Lecoeur	R	Phased array system featuring analog transmitters for arbitrary waveform generation.	[64]
SARUS	U	Large scale experimental system with 1024 TX/RX channels.	[65]
ULA-OP	U	Open research platform. Portable system with PC interface.	[66]
Ultrasonix	C	Clinical system with add-on HW unit for access to raw data. Programmable HW beamformer.	[67]
Verasonics	R	Flexible platform with real-time access to raw channel data. Beamforming done in SW using pixel-based techniques.	[68]
Visualsonics	R	High frequency system (up to 70 MHz) for small animal studies.	[69]
Zonare	C	Basic research functions. Allows access to RF data (pre BF).	[70]

[a] R = research only systems; C = clinical systems; I = research interfaces for existing hardware; U = noncommercial research systems developed at universities. BF, beamformer; HW, hardware; RX, receive; SW, software; TX, transmit.

IV. QC OF RESEARCH SYSTEMS

The techniques outlined in Sec. II and III are of primary importance within a clinical setting, in which robust QC is mandatory for reasons such as patient safety and regulatory compliance. Within such an environment it is implicit that the ultrasound equipment has been cleared for medical use by the relevant regulatory bodies. In the USA, such clearance must be obtained from the FDA and includes numerous factors such as maximum acoustic output power, how information is displayed to the operator and how specifications such as measurement accuracy are obtained [58]. As a result, clinical ultrasound systems are subject to numerous prescribed limitations which may affect their utility for research applications such as the development of new transducer designs, beamforming methods and image reconstruction algorithms. In order to provide the additional functionality required for such applications a number of ultrasound research platforms have been developed. These include clinical systems featuring a separate 'research mode' as well as dedicated research-only systems. As they are not subject to the same certification requirements as clinical devices these may provide the additional flexibility required for research applications, such as arbitrary manipulation of transmit/ receive settings or access to raw ultrasound data.

While QC may not be explicitly mandated for research systems it is still of great importance for the integrity of research studies and to ensure that results are robust and reproducible. Despite this, QC is rarely discussed in the context of research-only ultrasound systems in the literature

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

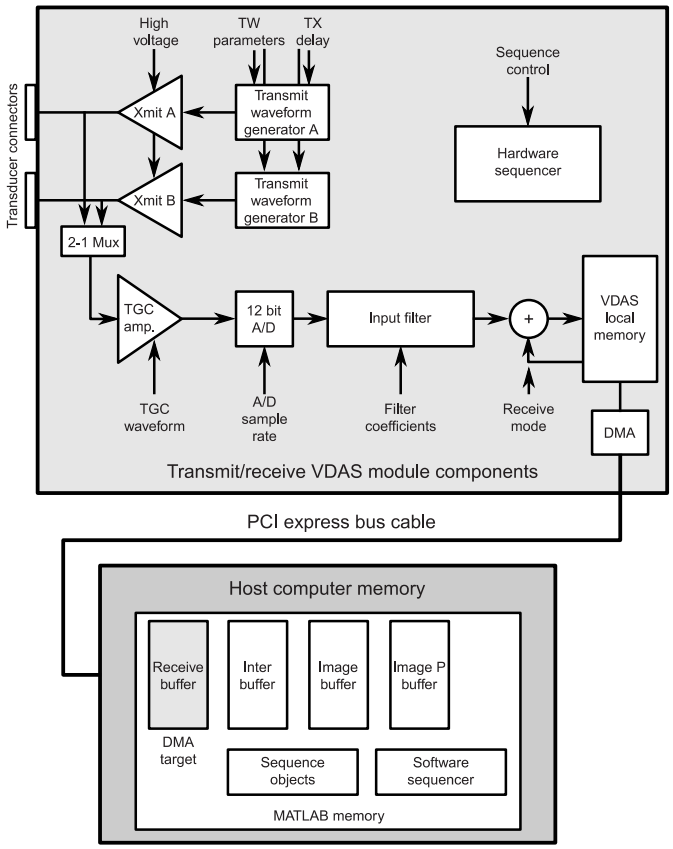

Fig. 11. Verasonics ultrasound research system block diagram. The upper box represents the data acquisition (hardware) system which is fitted with conventional probe connectors for input/output. The lower box represents memory within the host PC system. direct memory access (DMA); Verasonics data-acquisition system (VDAS); transmit (Xmit). Based on [60].

other than with respect to application-specific metrics, such as using cross-correlation coefficients as an indication of the robustness of displacement calculations for shear wave elastography [59]. In the following sections major ultrasound research platforms are briefly reviewed, a high-level summary of a particular research-only system (*Verasonics*) is introduced, and the implications for application of QC to such research systems are discussed.

A. Review of Ultrasound Research Systems

A summary of major ultrasound systems which have been used for research applications is given in Table III. These may be broadly categorized as clinically-approved scanners (with varying degrees of research functionality), commercial research-only systems, research platforms developed in academic labs, and other niche systems targeting particular research needs such as modularity or transmit waveform generation. For those systems which are clinically approved, QC may initially be applied under the clinical mode using the methods discussed in Sec. II and III, which will account for many common failure modes such as probe damage. For research-only systems (and research-specific parts of clinical systems) such methods may or may not be possible, depending upon the application under consideration. In addition, the lower-level control and data access afforded by research systems necessitates consideration of many additional factors. To illustrate this point in the following section the structure of the *Verasonics* research system is introduced. This system was selected as it is (i) research only; (ii) offers powerful research capabilities (full real-time access to raw channel data) and (iii) commercially available and thus employed by several research groups in the literature.

B. The Verasonics Research System

A simplified block diagram of the Verasonics ultrasound research platform is shown in Fig. 11. The upper box comprises the 'VDAS' data acquisition hardware, the lower represents the relevant memory locations on a host computer system. The hardware unit contains many of the components of a typical ultrasound scanner including waveform generators and amplifiers for transmission of pulses, and amplification, A/D conversion and filtering for receive. Unlike a typical clinical scanner each of these components may be directly controlled by the user over a wide range of parameters – for example high intensity transmit pulses may be generated (in excess of FDA diagnostic limits) for therapeutic applications, or transmission disabled entirely for passive imaging – allowing extensive control over transmit, receive, signal processing and data transfer to the host [71]. Notably, the hardware unit was designed to facilitate access to raw RF data so does not contain a beamformer; as such, image reconstruction must be conducted entirely in software. In order to facilitate this, the hardware unit is linked to the host PC using a high speed bus: raw data are transferred directly to memory of the host PC where they may be reconstructed in real-time using pixel-based methods [72] for conventional imaging, passed to functions developed by the researcher for immediate processing, or recorded to disk for later analysis. The system interfaces to the popular MATLAB® programming environment which allows development of setup scripts and custom interfaces for a particular task. The system has been used for a range of applications in the literature including development of custom transducer arrays [73], shear wave elastography [59] and passive monitoring of cavitation [74].

C. Application of QC to Research Systems

In this section the application of QC to research systems is discussed using the example of the Verasonics architecture as introduced above. While the Verasonics platform is used as an example here, it should be noted that other systems such as the Antares research platform, ULA-OP and Ultrasonix may also be configured to write data to a MATLAB® programming environment and QC methods applied in a similar manner. To remain within the scope of this paper it is assumed that QC is primarily to be conducted using B-mode imaging (e.g. using a conventional imaging probe). This assumption is reasonable as the Verasonics system is plug-and-play compatible with a range of readily-available imaging arrays [75], so testing in such a manner would require minimal additional effort and should be sufficient to account for a broad range of failure modes or changes in system performance. Appropriate QC for other advanced research goals is application-specific but could include similar methods to those outlined below, such as use

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2016.2619622, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

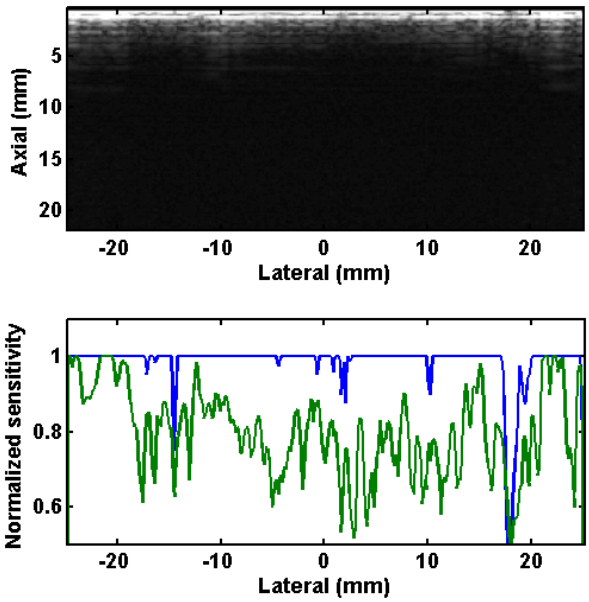

Fig. 12. Example of automated QC image analysis using the Verasonics system. Top: B-mode image showing reverberation pattern in air (relative uniformity test) for a linear array. Bottom: Plot of normalized brightness vs. lateral pixel location at two different axial pixel locations (0.6mm (blue) and 0.9mm (green)) shows uneven sensitivity and suggests possible element damage at around +18mm lateral.

of imaging metrics for evaluation of new beamforming algorithms. The discussion below follows the QC methods discussed in Sec. II.

1) Level 1– Performance testing by visual inspection

Level 1 routine testing may be conducted in a similar manner to clinical scanners. Relative uniformity may be tested for a given probe by examining the reverberation pattern in air using one of the basic B-mode imaging scripts supplied with the system. As with clinical scanners the gain and compression may be adjusted to improve contrast of the images. Since these parameters are set in software they may be pre-set for regular testing by creating a customized imaging script, which would also ensure that the various other parameters which may be set on the research system (Fig. 11) remain constant for comparison of results over time. The relative noise test may be conducted in a similar manner by adjustment of gain and TGC values as described previously.

Results of either test may be compared over time by simply saving images to disk for visual assessment, or using an automated process. Automated analysis may be achieved using standard programming methods in the MATLAB® programming environment: for example, relative uniformity images may be analyzed by extracting image brightness as a function of lateral position in a region of interest containing the reverberation pattern (Fig. 12). Regular testing in such a manner would provide quick verification of system performance and reveal issues such as probe damage or other issues in the transmit/receive signal chain or software environment.

2) Level 2&3– Performance and acceptance testing by use of ultrasound phantoms and image analysis software

Level 2 and 3 testing typically involves use of phantoms and analysis of images in software. While not explicitly required, the use of standard commercially-available phantoms may be preferable for robustness and to facilitate easier comparison of results between different probes, systems and research groups. As per Level 1 testing customized imaging scripts should be created for each type of test by modifying the parameters illustrated in Fig 11. The tests described in Sec. II may be applied both for testing the performance of 'known' probes (i.e. those for which imaging scripts are provided with the Verasonics system) as well as for verification of probe parameters (element spacing, frequency response etc.) for custom transducers. Tests such as distance accuracy, slice thickness and dynamic range could all be conducted in the same manner as described for clinical systems, and routines written for their analysis in the MATLAB® environment. Testing in such a manner would allow verification of system performance for known transducers as well as providing a framework for optimization of imaging performance (using well established techniques) for custom transducer arrays in research applications. Finally, it should be noted that raw RF data are readily available on most research scanners, which could be used directly to perform QC rather than by imaging-based methods. This could improve the sensitivity of analysis by including the full dynamic range of the data as well as phase information, thus improving the performance of QC and expanding the scope to include applications in which conventional B-mode imaging is insufficient.

V. Conclusions

QC of ultrasound equipment is not as well established clinically as for the other imaging modalities. It is however advised by ultrasound and medical professional organizations and recently some accrediting agencies have made mandatory the documentation of ultrasound QC. QC is important for detecting equipment failure and for improving image quality. We have discussed a number of subjective and objective methods for QC of US imaging systems with focus on breast imaging and biopsy. These tests include: accuracy in distance measurement, contrast and spatial resolution, uniformity and sensitivity for B-mode imaging. We have discussed these methods in details with the goal of making it easier to implement them and also for encouraging the development of QC methods suitable for clinical US imaging systems.

References

[1] E. B. Mendelson, J. K. Baum, W. A. Berg, C. R. Merritt, E. Rubin, "Breast Imaging Reporting and Data System – Ultrasound ACR BI-RADS®-US,"(first edition), Reston, VA, American College of Radiology, 2003.

[2] A. T. Stavros, "Breast Ultrasound Equipment Requirements," in *Breast Ultrasound*, Philadelphia, Pa: Lippincott Williams & Wilkins, 2004, Ch. 1, pp. 16-41.

[3] AIUM, "Methods for measuring performance of pulse-echo ultrasound imaging equipment, part II: digital methods, stage 1," American Institute of Ultrasound in Medicine, 1995.

[4] AIUM, "Technical Standards Committee. Quality control manual for gray-scale ultrasound scanners – stage 2," 1995.

[5] AIUM, "Routine Quality Assurance for Diagnostic Ultrasound Equipment," 2008.

[6] A. Goodsitt, P. L. Carson, S. Witt, D. L. Hykes, J. M. Jr Kofler, "Real time B-mode ultrasound quality control test procedures: report of AAPM task group n°1," *Med Phys.*, vol. 25, no.8, pp. 1385-1406.

[7] ACR, "Breast Ultrasound Accreditation Program Requirements," 2016.

[8] C. Kollmann, C. deKorte, N. J. Dudley, N. Gritzmann, K. Martin, D. H. Evans; EFSUMB Technical Quality Assurance Group--US-TQA/B. "Guideline for Technical Quality Assurance (TQA) of ultrasound devices (B-Mode)--version 1.0 (July 2012): EFSUMB Technical Quality Assurance Group--US-TQA/B," *Ultraschall Med.*, vol. 33, no. 6, pp. 544-549, Dec. 2012.

[9] IPEM, Quality Assurance of Ultrasound Imaging Systems, 2010.

[10] J. M. Thijssen, G. Weijers, C. L. de Korte, "Objective performance testing and quality assurance of medical ultrasound equipment," *Ultrasound Med. Biol.*, vol. 33, no.3, pp. 460–471, 2007.

[11] N. M. Gibson, N. J. Dudley, K. Griffith, "A Computerised Ultrasound Quality Control Testing System," *Ultrasound Med. Biol*., vol. 27, no. 12, pp. 1697–1711, 2001.

[12] J. E. Browne, A. J. Watson, N. M. Gibson, N. J. Dudley, A. T. Elliott, "Objective measurements of image quality," *Ultrasound Med. Biol*., vol. 30, no. 2, pp. 229–237, 2004.

[13] International Electrotechnical Commission (IEC). IEC 62736 Ed. 1.0 Quality Control of Diagnostic Medical Ultrasound Systems. Draft Geneva. 2011.

[14] International Electrotechnical Commission (IEC). IEC 61391-1 Edition 1.0. Ultrasonics – Pulse-echo scanners – Part 1: Techniques for calibrating spatial measurement systems and measurement of system point-spread function response, 2006.

[15] International Electrotechnical Commission (IEC). IEC 61391-2 Edition1.0. Ultrasonics – Pulse-echo scanners – Part 2: Measurement of maximum depth of penetration and local dynamic range, 2010.

[16] E. Samei, A. Badano, D. Chakraborty, K. Compton, et al. "Assessment of display performance for medical imaging systems: Executive summary of AAPM 18 report," *Med Phys.*, vol. 32, no. 4, pp.1205-1225, 2005.

[17] N. J. Hangiandreou, S. F. Stekel, D. J. Tradup, K. R. Gorny, D. M. King, "Four-year experience with a clinical ultrasound Quality Control program," *Ultrasound Med. Biol*., vol. 37, no. 8, pp. 1350–1357, 2011.

[18] O. Sipila, V. Mannila, E. Vartiainen, "Quality assurance in diagnostic ultrasound," *Eur. J. Radiol*., vol. 80, no. 2, pp. 519–525, 2011.

[19] T. Evans, P. Hoskins, "Quality Assurance", in *Diagnostic Ultrasound Physics and Equipment*, P. Hoskins, K. Martin, A. Thrush, Eds., Cambridge: Cambridge University Press, 2010, Ch. 11, pp. 142-154.

[20] J. E. Browne, K. V. Ramnairne KV, A. J. Watson, P. R. Hoskins, "Assessment of the Acoustic Properties of common tissue-mimicking test objects," *Ultrasound Med. Biol.*, vol. 29, no.7, pp 1053 - 1060, 2003.

[21] A. Goldstein, "The effect of acoustic velocity on phantom measurements," *Ultrasound Med. Biol.*,, vol 26, no. 7, pp. 1133 -1143, 2000.

[22] A. Goldstein, "Slice Thickness Measurements," *J. Ultrasound Med*., vol. 7, no. 9, pp. 487–498, 1988.

[23] B. Richard, "Test Object for Measurement of Section Thickness at US," *Radiology*, vol. 211, no.1, pp. 279–282, 1999.

[24] F. P. Branca, S. A. Sciuto, A. Scorza, "Comparative evaluation of ultrasound scanner accuracy in distance measurement," *Rev. Sci. Instrum.*, vol. 83, no. 10, pp. 105103, 2012.

[25] A. Scorza, S. Conforto, C. D'Anna, S. A. Sciuto, "A comparative study on the influence of probe placement on quality assurance measurements in B-mode Ultrasound by means of ultrasound phantoms," *Open Biomed. Eng. J.*, vol. 9, pp. 164-178, 2015.

[26] M. L. Skolnick, "Estimation of ultrasound beam width in the elevation (section thickness) plane," *Radiology*, vol. 180, no. 1, pp. 286–288, 1991.

[27] A. Scorza, "A novel method for automatic evaluation of the effective dynamic range of medical ultrasound scanners," *ECIFMBE 2008, IFMBE Proceedings*, vol. 22, Berlin Heidelberg: Springer-Verlag, 2009, pp. 1607–1611.

[28] J. A. Jensen, "Medical ultrasound imaging," *Prog. Biophys. Mol. Biol*., vol. 93, no. 1-3, pp. 153-165, 2007.

[29] A. Scorza, G. Lupi, S. A. Sciuto, F. Bini, F. Marinozzi, "A novel approach to a phantom based method for maximum depth of penetration measurement in diagnostic ultrasound: a preliminary study," in *IEEE International Symposium on Medical Measurements and Applications (MEMEA 2015) Proc*, Torino, Italy, pp.369-374, 2015.

[30] K. R. Gorny, D. J. Tradup, N. J. Hangiandreou, "Implementation and validation of three automated methods for measuring ultrasound maximum depth of penetration: application to ultrasound quality control," *Med. Phys.*, vol. 32, no. 8, pp. 2615–28, 2005.

[31] M. Martensson, M. Olsson, B. Segall, A. G. Fraser, R. Winter, L. A. Brodin, "High incidence of defective ultrasound transducers in use in routine clinical practice," *Eur. J. Echocardiogr.*, vol. 10, no. 3, pp. 389–394, 2009.

[32] N. Hangiandreou, Ultrasound Quality Control: A Practical Overview of Tests, Methods, and ACR Accreditation Requirements, AAPM Annual Meeting, 2013.

[33] A. Scorza, G. Lupi, S. A. Sciuto, L. Battista, J. Galo, "A preliminary study on a method for objective uniformity assessment in diagnostic ultrasound imaging." In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC 2015) Proc*, Pisa, Italy, pp. 1628-1633, 2015

[34] A. Scorza, S. Conforto, G. Lupi and S. A. Sciuto "A texture analysis approach for objective uniformity evaluation in diagnostic ultrasound imaging: a preliminary study" in: *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015) Proc*, Milano, Italy, pp. 6317-6320, 2015.

[35] I. Schafhalter-Zoppoth, C. E. McCulloch, A. T. Gray, "Ultrasound visibility of needles for regional nerve block: an in vitro study," *Reg. Anesth. Pain. Med.*, vol. 29, no.5,pp. 480–9, 2004.

[36] S. Karstrup, J. Brøns, L. Morsel, N. Juul, P. von der Recke, "Optimal set-up for ultrasound guided punctures using new scanner applications: an in-vitro study," *Eur. J. Ultrasound.*, vol. 15, no. 1-2, pp. 77–84, 2002.

[37] K. J. Chin, A. Perlas, R. Brull, "Needle Visualization in Ultrasound-Guided Regional Anesthesia: Challenges and Solutions," *Reg. Anesth. Pain. Med.*, vol. 33, no. 6, pp 532–544, 2008

[38] T. Maecken, M. Zenz, T. Grau, "Ultrasound Characteristics of Needles for Regional Anesthesia," *Reg. Anesth. Pain Med.*, vol.32, no. 5, pp. 440-447, 2007.

[39] K. Nichols, L. B. Wright, T. Spencer, W. C. Culp, "Changes in ultrasonographic echogenicity and visibility of needles with changes in angles of insonation," *J. Vasc. Interv. Radiol.*, vol. 14, no. 12, pp.1553-1557, 2003.

[40] R. E. Hopkins, M. Bradley, "In-vitro visualization of biopsy needles with ultrasound: a comparative study of standard and echogenic needles using an ultrasound phantom," *Clin. Radiol*., vol. 56, no. 6, pp. 499-502, 2001.

[41] K. Nakagawa, T. Kamiya, K. Arakawa, S. Akiyama, K. Sakai, "Objective and subjective comparison of the visibility of three echogenic needles and a nonechogenic needle on older ultrasound devices," *Acta Anaesthesiol. Taiwanica*, vol 53, no. 1, pp. 1-6, 2015.

[42] S. Guo, A. Schwab, G. McLeod, G. Corner, S. Cochran, R. Eisma, R. Soames, "Echogenic regional anaesthesia needles: a comparison study in Thiel cadavers," *Ultrasound Med. Biol*. vol. 38, no. 4, pp. 702-707, 2012.

[43] H. Edgcombe, G. Hocking, "Sonographic identification of needle tip by specialists and novices: a blinded comparison of 5 regional block needles in fresh human cadavers," *Reg. Anesth. Pain Med.*, vol. 35, no. 2, pp. 207-211, 2010.

[44] S. Hebard, G. Hocking, K. Murray, "Two-dimensional mapping to assess direction and magnitude of needle tip error in ultrasound-guided regional anaesthesia." *Anaesth. Intensive Care* vol. 39, no. 6, pp. 1076-1081, 2011.

[45] S. Brekke, C. B. Ingul, S. A. Aase, H. G. Torp, "Increasing Frame Rate in Ultrasound Imaging by Temporal Morphing Using Tissue Doppler," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 53, no. 5, 2006.

[46] S. Park, S. R. Aglyamov, S. Y. Emelianov, "Elasticity Imaging Using Conventional and High-Frame Rate Ultrasound Imaging: Experimental Study," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 54, no. 11, 2007.

[47] L. Zhang, X. Xu, C. Hu, L. Sun, J. T. Yen, J. M. Cannata, K. K. Shung, "A High-Frequency High Frame Rate Duplex Ultrasound Linear Array Imaging System for Small Animal Imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 57, no. 7, pp. 1548–1557, 2010.

[48] F. P. M. Oliveira, J. M. R. S. Tavares, "Medical Image Registration: a Review," *Comput. Methods Biomech. Biomed. Engin.*, vol. 17, no. 2, pp:73-93, 2014.
[49] Optimizer 1425A Ultrasound Image Analyzer for Doppler and Grey Scale Scanners - User's Guide, Gammex, Inc. 2003.
[50] S. Balbis, T. Meloni, S. Tofani, F. Zenone, D. Nucera, C. Guiot, "Criteria and Scheduling of Quality Control of B-Mode and Doppler Ultrasonography Equipment," *J. Clin. Ultrasound*, vol. 40, no. 3, pp. 167-173, 2012
[51] S. Balbis, C. Musacchio, C. Guiot, R. Spagnolo, "US quality control in Italy: present and future," *Journal of Physics:Conference*, Series 279 pp. 012005, 2011.
[52] EN 61266:1995-04, Ultrasonics. Hand-held probe Doppler foetal heartbeat detectors. Performance requirements and methods of measurement and reporting.
[53] J.M. Thijssen, M.C. van Wijk, M.H.M. Cuypers, "Performance testing of medical echo/Doppler equipment," *Eur. J. Ultrasound*, vol. 15, no.3, pp. 151-164, 2002.
[54] J. E. Browne, A. J. Watson, P. R. Hoskins, A. T. Elliott, "Validation of a sensitivity performance index test protocol and evaluation of colour Doppler sensitivity for a range of ultrasound scanners," *Ultrasound Med. Biol.*, vol. 30, no. 11, pp. 1475–1483, 2004.
[55] H. Madjar, H.J. Prompeler, C. Del Favero, B.J. Hackeloer, J.B. Llull, "A new Doppler signal enhancing agent for flow assessment in breast lesions," *Eur. J. Ultrasound*, vol. 12, no. 2, pp. 123–130, 2000.
[56] I.A. Wright, N.D. Pugh, K. Lyons, D.J.T. Webster, R.E. Mansel, "Power Doppler in breast tumours: a comparison with conventional colour Doppler imaging," *Eur. J. Ultrasound*, vol. 7, no. 3, pp. 175–181, 1998.
[57] P. Milz, A. Lienemann, M. Kessler, M. Reiser, "Evaluation of breast lesions by power Doppler sonography," *Eur. Radiol.*, vol. 11, no. 4, pp 547-554, 2001.
[58] US FDA, "Guidance for Industry and FDA Staff Information for Manufacturers Seeking Marketing Clearance of Diagnostic Ultrasound Systems and Transducers," Document issued on September, vol. 9, 2008.
[59] M. Mehrmohammadi, P. Song, D. D. Meixner, R. T. Fazzio, S. Chen, J. F. Greenleaf, M. Fatemi, and A. Alizad, "Comb-Push Ultrasound Shear Elastography (CUSE) for Evaluation of Thyroid Nodules: Preliminary In Vivo Results," IEEE Transactions on Medical Imaging, vol. 34, no. 1, pp. 97–106, Jan. 2015.
[60] T. L. Szabo, Diagnostic Ultrasound Imaging: Inside Out. Academic Press, 2013.
[61] M. Ashfaq, S. S. Brunke, J. J. Dahl, H. Ermert, C. Hansen, and M. F. Insana, "An ultrasound research interface for a clinical system," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 53, no. 10, pp. 1759–1771, Oct. 2006.
[62] H. J. Hewener, H. J. Welsch, H. Fonfara, F. Motzki, and S. H. Tretbar, "Highly scalable and flexible FPGA based platform for advanced ultrasound research," in 2012 IEEE International Ultrasonics Symposium, 2012, pp. 2075–2080.
[63] L. Masotti, E. Biagi, M. Scabia, A. Acquafresca, R. Facchini, A. Ricci, and D. Bini, "FEMMINA real-time, radio-frequency echo-signal equipment for testing novel investigation methods," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 53, no. 10, pp. 1783–1795, Oct. 2006.
[64] P. Kruizinga, F. Mastik, N. De Jong, A. F. M. Van der Steen, and G. van Soest, "Ultrasound-guided photoacoustic image reconstruction," 2012, vol. 8223, pp. 822339.
[65] J. A. Jensen, H. Holten-Lund, R. T. Nielson, B. G. Tomov, M. B. Stuart, S. I. Nikolov, M. Hansen, and U. D. Larsen, "Performance of SARUS: A synthetic aperture real-time ultrasound system," in *Proc. IEEE Ultrasonics Symp.*, Oct. 2010, pp. 305–309.
[66] P. Tortoli, L. Bassi, E. Boni, A. Dallai, F. Guidi, and S. Ricci, "ULA-OP: an advanced open platform for ultrasound research," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 56, no. 10, pp. 2207–2216, Oct. 2009.
[67] T. Wilson, J. Zagzebski, T. Varghese, Q. Chen, and M. Rao, "The ultrasonix 500RP: A commercial ultrasound research interface," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 53, no. 10, pp. 1772–1782, Oct. 2006.
[68] P. J. Kaczkowski and R. E. Daigle, "The Verasonics ultrasound system as a pedagogic tool in teaching wave propagation, scattering, beamforming, and signal processing concepts in physics and engineering.," The Journal of the Acoustical Society of America, vol. 129, no. 4, pp. 2648–2648, 2011.
[69] F. S. Foster, J. Hossack, and S. L. Adamson, "Micro-ultrasound for preclinical imaging," Interface Focus, vol. 1, no. 4, pp. 576–601, Aug. 2011.
[70] L. Y. L. Mo, D. DeBusschere, W. Bai, D. Napolitano, A. Irish, S. Marschall, G. W. McLaughlin, Z. Yang, P. L. Carson, and J. B. Fowlkes, "P5C-6 Compact Ultrasound Scanner with Built-in Raw Data Acquisition Capabilities," in IEEE Ultrasonics Symposium, 2007, 2007, pp. 2259–2262.
[71] Verasonics, Inc., "A tool for education", [Online]. Avalable: http://verasonics.com/a-tool-for-education/
[72] R. E. Daigle, "Ultrasound imaging system with pixel oriented processing," US Patent 8287456, 2012.
[73] F. C. Meral and G. T. Clement, "128 Element ultrasound array for transcranial imaging," in 2010 IEEE International Ultrasonics Symposium, 2010, pp. 1984–1987.
[74] C. D. Arvanitis and N. McDannold, "Integrated ultrasound and magnetic resonance imaging for simultaneous temperature and cavitation monitoring during focused ultrasound therapies," Medical Physics, vol. 40, no. 11, p. 112901, Nov. 2013.
[75] Verasonics, Inc., "Transducer resources", [Online]. Avalable: http://verasonics.com/transducer-resources/