Ultrasound in Med. & Biol., Vol. 33, No. 3, pp. 460–471, 2007

0301-5629/07/$–see front matter

doi:10.1016/j.ultrasmedbio.2006.09.006

● *Original Contribution*

OBJECTIVE PERFORMANCE TESTING AND QUALITY ASSURANCE OF MEDICAL ULTRASOUND EQUIPMENT

JOHAN M. THIJSSEN, GERT WEIJERS, and CHRIS L. DE KORTE
Clinical Physics Laboratory, University Children's Hospital, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

(*Received* 24 *May* 2006; *revised* 4 *September* 2006; *in final form* 19 *September* 2006)

Abstract—There is an urgent need for a measurement protocol and software analysis for objective testing of the imaging performance of medical ultrasound equipment from a user's point of view. Methods for testing of imaging performance were developed. Simple test objects were used, which have a long life expectancy. First, the elevational focus (slice thickness) of the transducer was estimated and the in-plane transmit focus was positioned at the same depth. Next, the postprocessing look-up-table (LUT) was measured and linearized. The tests performed were echo level dynamic range (dB), contrast resolution (*i.e.*, gamma of display, number of gray levels/dB) and sensitivity, overall system sensitivity, lateral sensitivity profile, dead zone, spatial resolution and geometric conformity of display. The concept of a computational observer was used to define the lesion signal-to-noise ratio, SNR_L (or Mahalanobis distance), as a measure for contrast sensitivity. All the measurements were made using digitized images and quantified by objective means, *i.e.*, by image analysis. The whole performance measurement protocol, as well as the quantitative measurements, have been implemented in software. An extensive data-base browser was implemented from which analysis of the images can be started and reports generated. These reports contain all the information about the measurements, such as graphs, images and numbers. The approach of calibrating the gamma by using a linearized LUT was validated by processing simultaneously acquired rf data. The contrast resolution and echo level of the rf data had to be compressed by a factor of two and amplified by a gain factor corresponding to 12 dB. This resulted in contrast curves that were practically identical to those obtained from DICOM image data. The effects of changing the transducer center frequency on the spatial resolution and contrast sensitivity were estimated to illustrate the practical usefulness of the developed approach of quality assurance by measuring objective performance characteristics. The developed methods might be considered as a minimum set of objective quality assurance measures. This set might be used to predict clinical performance of medical ultrasound equipment, taking into account the performance at a unique point in space *i.e.*, the coinciding depths of the elevation and in-plane (azimuth) foci. Furthermore, it should be investigated whether the approach might be used to compare objectively various brands of equipment and to evaluate the performance specifications given by the manufacturer. Last but not least, the developed approach can be used to monitor, in a hospital environment, the medical ultrasound equipment during its life cycle. The software package may be viewed and downloaded at the website http://www.qa4us.eu. (E-mail: j.thijssen@cukz.umcn.nl)

Key Words: **Computational observer, Contrast resolution, Echography, Image quality, Medical ultrasound, Objective assessment, Performance testing, Quality assurance, Spatial resolution, QA4US®.**

INTRODUCTION

The first efforts to develop performance methods for medical ultrasound equipment by various national and international committees date back 30 y [American Institute of Ultrasound in Medicine (AIUM 1974), the American Association of Physicists in Medicine (AAPM; Carson and Zagzebski 1977) and the International Electrotechnical Commission (IEC; Hill 1977)]. Progress has been slow for various reasons. Evolution of equipment features and performance is still very rapid and some ambiguity seems to be present within these committees in defining a clear endpoint; should it be a technical standard or a quality assessment from the user's point of view? Another limitation of the methods involved in quality standards, so far, has been the involvement of subjective assessments in case of image quality. This limitation also holds for a recent AAPM report (Goodsitt

Address correspondence to: Professor Johan M. Thijssen, PhD, 833-Clinical Physics Laboratory, University Children's Hospital, Radboud University Nijmegen Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. E-mail: j.thijssen@cukz.umcn.nl

et al. 1998), although it has evidently been based on the endpoint of quality of use, *i.e.*, imaging performance as well as for the paper by Dudley et al. (2001); see also Gibson et al. (2001). Although the recent paper by Browne et al. (2004) is based on objective assessment, a serious limitation is the use of gray level rather than relative echo level (in dB), as was proposed by Thijssen et al. (2002) and van Wijk and Thijssen (2002), which is incorporated also in new IEC standards being developed. Therefore, the dB scale is used in this paper for the estimation of most of the quality measures: overall dynamic range, contrast resolution, contrast sensitivity, spatial resolution and overall system sensitivity.

In this article, the approach is followed that, rather than by a subjective assessment, the performance measurements are carried out on echographic images, which are digitally acquired in a computer and recalibrated by using the postprocessing look-up table (LUT) and by measuring the gray level of scattering objects with a known contrast in a tissue-mimicking phantom. This allows for an adequate recalibration of gray levels to relative echo levels (dB) and for an objective assessment of image quality by using the concept of a computational observer (Lopez et al. 1992). Image quality is assessed by measuring the lesion signal-to-noise ratio (SNR_L; see Smith et al. 1983), which is to be considered as a contrast sensitivity measure. Other measures related to contrast are the overall echo level dynamic range and the contrast resolution.

The authors are of the opinion that quality assurance measurements should be made with a unique and reproducible setting of the elevation, transmit focus of a transducer. The depth of this elevation focus was chosen to coincide with the azimuth focus because it can be expected that with this setting the performance of a transducer is optimal over the depth-of-focus range. Furthermore, for a minimum set of performance measures, this single focus is considered to be sufficient.

New methods of assessment of quality and performance characteristics have been added to those published in the two aforementioned papers by the principal author (Thijssen et al. 2002 and van Wijk and Thijssen 2002). One test was omitted in the present paper, this was the assessment of the tissue-to-clutter ratio (TCR) of anechoic lesions. The reason for this is that the observed TCR was relatively high and did not change much with depth (van Wijk and Thijssen 2002). Ideally, this parameter should be measured with spherical cysts rather than with cylindrical bars, as are present in commercially-available phantoms.

The validity of the approach of using recalibrated echo levels (dB) was investigated by using simultaneously made radio-frequency signal acquisitions as a reference. The methods are illustrated by experiments regarding an assessment of image quality in relation to the choice of ultrasound frequency/bandwidth of the system.

TEST OBJECTS

The performance measurements were carried out by using tissue-mimicking phantoms made of urethane rubber base material (ATS Labs Inc., Bridgeport, CT, USA). The backscattering and the attenuation of the base material are similar to those of parenchymal soft tissues (speed of sound 1540 m/s; attenuation coefficient 0.3 to 0.5 dB/cm.MHz; backscattering 1 to 4.10^{-4}/m.sr). The speed of sound of the phantom at room temperature (c = 1432 m/s) is, however, 7% lower than in the human body. The manufacturer has adapted the geometry in the axial direction so as to prevent image distortion at room temperature. Nevertheless, the focusing of array transducers might be somewhat disturbed, leading to an overestimation (*i.e.*, worsening) of the resolution measurements (*i.e.*, of the size of the in-plane point-spread function, *e.g.*, Chen and Zagzebski 2004 and Goldstein 2004) and of the horizontal distance measurements (Goldstein 2002). In the authors' experience, the long preservation time of the phantom far outweighs the minor, but fixed, effects on resolution and geometrical conformity.

The imaging performance testing was carried out by using the ATS model 539 phantom for low frequency transducers, and ATS model 550 phantom for high frequency transducers. The main objects within the phantom to be used in the assessment scheme are thin wires in a cross-like pattern and cylindrical objects of known (nominal) scattering contrast compared with the surrounding material (Fig. 1). The nominal contrast values of these solid lesions with respect to the background gray level are specified by the manufacturer in decibels (dB), for the frequency range 2.25 to 7.5 MHz (ATS model 539) and ≥7.5 MHz (ATS model 550).

The measurement of the position and width of the elevation focus was made with the slice thickness phantom (ATS model 538N).

EQUIPMENT AND EXPERIMENTS

Equipment

The experiments were performed with a SONOS 7500 (Philips Medical Systems, Andover, MA, USA). A linear array transducer (3 to 11 MHz) was used for most of the measurements for illustrating this paper.

This system is provided with a custom-designed raw data output of scan lines in radio frequency format (16 bits rf data). Further outputs available are a standard analog video (PAL) format and a digital (DICOM) format output. The measurements taken from the analog video output were acquired by using a frame grabber