

Enhancing reproducibility of ultrasonic measurements by new users

Manojit Pramanik^a, Madhumita Gupta^b, Kajoli Banerjee Krishnan^{*c}

^aDepartment of Electrical Engineering, Indian Institute of Science, Bangalore, India;

^bGE Healthcare, Bangalore, India; ^cGE Global Research, Bangalore, India

ABSTRACT

Perception of operator influences ultrasound image acquisition and processing. Lower costs are attracting new users to medical ultrasound. Anticipating an increase in this trend, we conducted a study to quantify the variability in ultrasonic measurements made by novice users and identify methods to reduce it. We designed a protocol with four presets and trained four new users to scan and manually measure the head circumference of a fetal phantom with an ultrasound scanner. In the first phase, the users followed this protocol in seven distinct sessions. They then received feedback on the quality of the scans from an expert. In the second phase, two of the users repeated the entire protocol aided by visual cues provided to them during scanning. We performed off-line measurements on all the images using a fully automated algorithm capable of measuring the head circumference from fetal phantom images. The ground truth (198.1 ± 1.6 mm) was based on sixteen scans and measurements made by an expert. Our analysis shows that: (1) the inter-observer variability of manual measurements was 5.5 mm, whereas the inter-observer variability of automated measurements was only 0.6 mm in the first phase (2) consistency of image appearance improved and mean manual measurements was 4-5 mm closer to the ground truth in the second phase (3) automated measurements were more precise, accurate and less sensitive to different presets compared to manual measurements in both phases. Our results show that visual aids and automation can bring more reproducibility to ultrasonic measurements made by new users. For the **keywords**, select up to 8 key terms for a search on your manuscript's subject.

Keywords: Ultrasound, measurement, reproducibility, phantom, algorithm, automation, new user

1. INTRODUCTION

Ultrasonic measurements from B-mode images are known to be both patient and operator dependent and influenced by (1) perceptual image quality, (2) access to and acquisition of the clinically prescribed 2-d scan plane, and (3) adherence to clinical guidelines in the process of measurement¹⁻². In a typical ultrasound scan, the user roves the probe over the anatomy of interest until the image representing the scan plane of interest appears on the monitor, freezes the image, adjusts the controls (such as gain) on the keyboard for visual acuity, and uses the mouse-like caliper to make a measurement. While much of the imaging is performed by well-trained sonographers, lower cost is nudging ultrasound towards the point-of-care for use by minimally skilled users such as primary care physicians³.

Estimation of gestational age from fetal biometrics is an important aspect of obstetric ultrasound. Our group has previously reported results from algorithms to automate fetal biometric measurements and scanning with retrospective images acquired by trained experts⁴⁻⁷.

Perception plays a significant role in interpreting images⁸. So, a radiologist feels 'pleased' by the appearance of an image based on its contrast, brightness, level of smoothness, uniformity of illumination and other qualitative attributes. Though perception is most often 'subjective', they are strongly related to the goal of imaging - what is being imaged, detected, measured and assessed.

*kajoli.krishnan@ge.com; phone 91 80 40122107; fax 91 80 28412111; ge.geglobalresearch.com

In this study we seek to explore the role of perception in the images acquired and measurements made by novice users of Ultrasound, influence of feedback and practice; influence of visual aids during scanning; and finally making an assessment of the efficacy of one of our automation algorithms for their use. We have first taught four novice users to obtain a 2D section of the head of a fetal phantom with a general imaging scanner. These minimally trained users have then scanned the fetal head on seven sessions for four different presets over a period of three months. Two of the novice users then scanned on another seven sessions with a visual feedback while scanning. We have analyzed and assessed the variability of automated measurement of fetal head circumference (HC) along with manual measurements performed by the new users in comparison with the ground truth as measured by an ‘expert’. Section 2 of this paper describes our experimental procedure. Analysis framework and results are presented in Section 3. Implications of the results and our conclusions are summarized in Section 4.

Studies based on obstetric patient scans have reported intra-observer and inter-observer variability of manual measurements by experts⁹ and also suggested that automated measurements are more consistent than manual measurements by multiple experts¹⁰. Image acquisition by experts in these studies has also implied that adherence to guidelines for scan plane and perceptual quality of images meets an acceptable standard. This study, on the other hand focuses on images acquired by novice users of ultrasound. Algorithmic optimization of the image quality is not invoked. It eliminates the influence of patient by employing a fetal phantom unlike the expert based studies in which statistically significant sample sizes obviate the need to explicitly consider the intrinsic patient dependence of ultrasound. We report intra-observer and inter-observer variability of manual measurement of head circumference, compare them with measurements made by an automated algorithm and benchmark all of them against the ‘ground truth’ based on images acquired and measured by an expert. Image acquisition by non-experts implies that adherence to guidelines for scan plane and perceptual quality of images may not always meet an acceptable standard. We provide expert feedback and visual assistance in the form of a reference image to the users and repeat the entire protocol. To the best of our knowledge, this is a first attempt to quantify the novice-user dependence of ultrasonic measurements, and simultaneously assess the usefulness of visual cues and automated measurements for them.

2. EXPERIMENTAL PROCEDURE

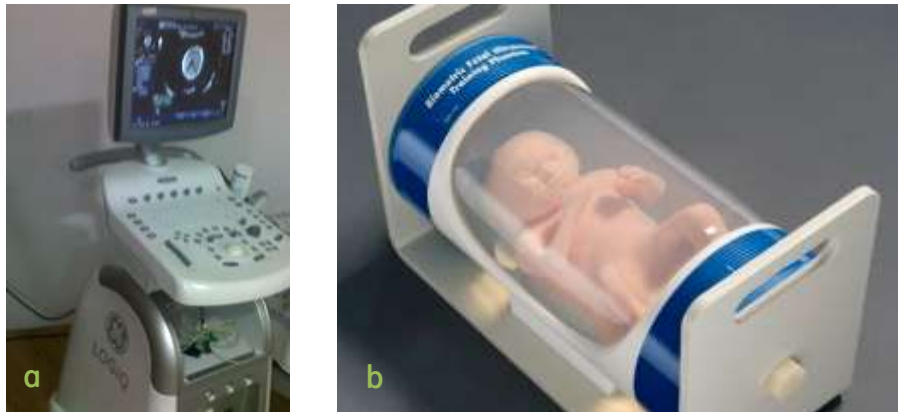


Figure 1. (a) GE LOGIQ P3 ultrasound scanner (b) CIRS Model 068 fetal phantom

Our study was based on scanning a fetal phantom (CIRS, Model 068) on a GE LOGIQ P3 ultrasound scanner (GE Healthcare, Bangalore). The phantom and the scanner are depicted in figures 1. The experimental protocol consisted of two phases. The first was a Training phase during which an application specialist (henceforth referred to as ‘expert’) trained four novice users (NV1, NV2, NV3, NV4) to obtain a 2-D section of the head of the fetal phantom appropriate for measurement of the head circumference (HC) and measure the HC with LOGIQ P3, a general imaging scanner. These minimally trained users then scanned the fetal head during seven distinct sessions for four different presets over a

period of three months. The user had the option to alter the brightness (gain) and contrast (dynamic range or DR) of the image. At the end of the first phase, the expert gave feedback to these users on the appropriateness of the scan plane, image quality and measurement on all their scans. In the second phase, two of the novice users (NV3 and NV4) repeated the procedure with an image representing a scan plane acquired by an expert placed in front of them during scanning. An automated algorithm was also employed to measure HC for all the images that were acquired by the novice users.

Three parameters of the standard obstetric preset on the scanner were modified for the purposes of this experiment. Based on inputs from the expert, two focal positions (f), one within the head region and one beyond the distal boundary of the head were chosen in combination with two sub-optimal settings of gain and DR to constitute a total of four presets. The presets were designed such that the users had sufficient scope to adjust the brightness and contrast of the image either during acquisition or after freezing prior to the measurement step. All the parameter values were invisible to the user to eliminate the possibility of additional bias during image acquisition. *Probe*: Convex 4C; *Fixed parameters*: Time gain compensation (TGC) pods – all centered, Transmit frequency = 4 MHz, Imaging depth = 13 cm, Frame rate = 35 Hz. *Preset 1*: Gain = 98, DR = 54 dB, f = 6 cm. *Preset 2*: Gain = 98, DR = 54 dB, f = 12 cm. *Preset 3*: Gain = 40, DR = 96 dB, f = 6 cm. *Preset 4*: Gain = 40, DR = 96 dB, f = 12 cm.

For this study, we applied our previously published algorithm to automatically detect and segment fetal head and measure its circumference from B-mode ultrasound images⁵. The parameters of this algorithm are based on approximately 400 images from three different scanners acquired by a trained gynecologist, expert sonographer, and trained paramedics, respectively and validated by 30 images from Voluson E8. We adapted this algorithm for application to the head of the fetal phantom.

3. RESULTS

Figure 2(a) shows a representative head scan by the expert. Figure 2(b) shows an example of a head scan by a novice user and manually drawn HC. Figure 2(c) shows the same image as (b) but measured by the automated method. The expert scanned a total of 16 times (4 presets X 4 sessions). Each of the four novices scanned a total number of 28 times (4 presets X 7 sessions). The variance associated with both image quality (as characterized by brightness and contrast) and manual measurement of HC is much lower for the expert as shown in figure 3, 4 and 5 respectively. We have considered mean HC measurement by expert as the ground truth in our analysis.



Figure 2. Fetal phantom head scan by (a) expert (b) novice user manual contour in yellow (c) automated contour in red

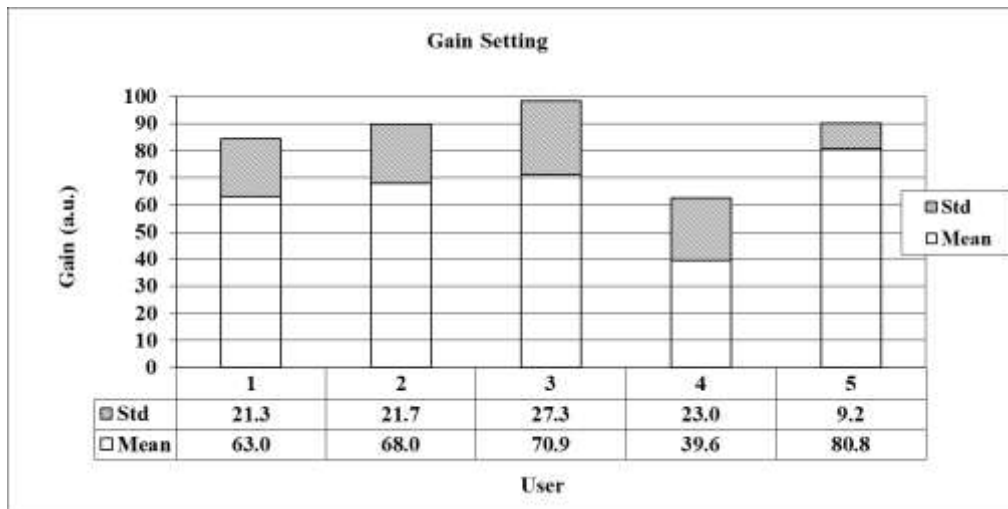


Figure 3. Mean and variability of brightness in images acquired by new users (1-4) and expert (5).

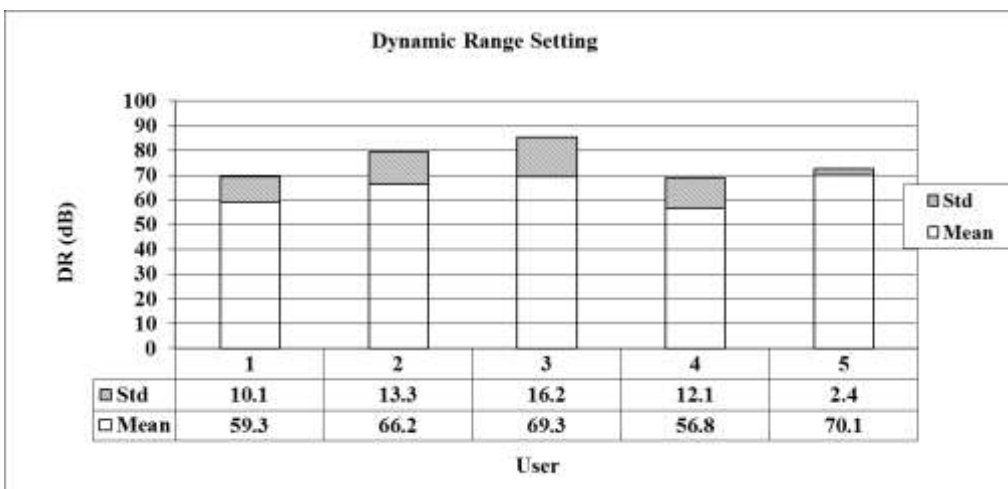


Figure 4. Mean and variability of contrast in images acquired by new users (1-4) and expert (5).

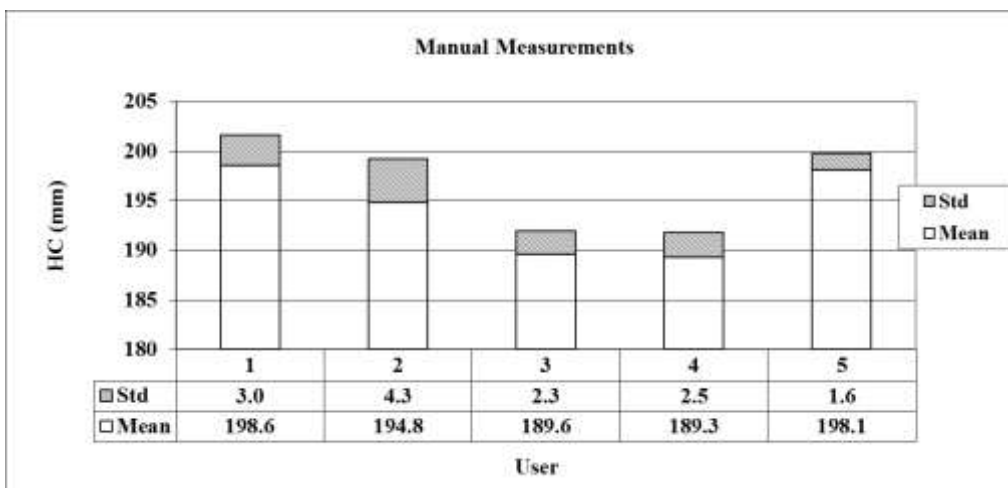


Figure 5. Mean and variability of manual measurements of HC in images acquired by new users (1-4) and expert (5).

Statistical analysis was done using Minitab¹¹ Release12. An Anderson Darling Normality Test with 95% confidence level performed on each of the five sets of manual measurements show that all of them follow a normal distribution with p-values: 0.922 (NV1), 0.242 (NV2), 0.870 (NV3), 0.934 (NV4), 0.081 (Expert). It thus becomes possible to calculate the intra-observer variability for the novices as the mean of $\Sigma(\mu_i - \mu_j) \quad i < j, 1 \leq i, j \leq 4$; μ 's are the mean manual measurements. An Anderson Darling Normality Test with 95% confidence level performed on automated measurements based on images acquired by novice users are also normally distributed; p-values: 0.056 (NV1), 0.562 (NV2), 0.470 (NV3), 0.337 (NV4). The mean intra-observer variability for manual and automated method is 3mm and 2.7mm while the inter-observer variability is 5.5mm and 0.6mm. Automation clearly makes the act of measurement more precise by radically improving upon the inter-observer variability in manual measurement by novice users. In fact, it is more precise (± 0.6 mm) than manual measurement by the expert (± 1.6 mm).

Two of the novice users (NV3 and NV4) received a composite rating on the images scanned by them from the expert at the end of training phase. The rating (good, acceptable or bad) was based on image quality (brightness, contrast), appropriateness of the scan plane and placement of the measurement contour. There were appreciable number of badly rated images (32% and 18%) and no or very few good images (0% and 11%).

After receiving the expert feedback, these two novice users repeated the entire protocol (4 presets X 7 times each) with the difference that they now had a sample image from an expert scan available for reference. 28 manual measurements each made by novices 3 and 4 before and after feedback are shown in figure 6. The trial numbers also mark the progression in time. The manual measurements clearly underestimate the HC during the Training phase (left side of the vertical bar) compared to the ground truth. This could occur due to the quality of the image acquired or the manual drawing or both. After the feedback, as time progresses, manual measurements move closer to the expert measurement. Likewise, the expert rating on image quality improves sharply (11% and 0% bad images).

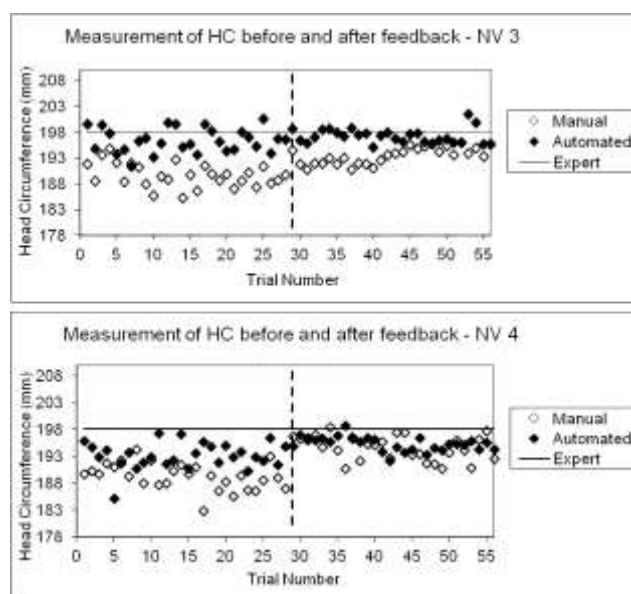


Figure 6. HC measurements by novice users 3 (Top) and 4 (Bottom). The dashed vertical bar separates the measurements before and after feedback

4. CONCLUSIONS

In summary, automated algorithm is more precise at measuring the HC than both novice users and the expert. Even during the training phase, automated method enables a degree of user-independence to measurements. Mean inter-observer variability is the lowest for automated method (0.6 mm) and well within 1 SD of 0.1 ± 5.6 mm reported for HC measurement of real patient scan in a clinical setting⁹. The fidelity of the scan plane to clinical guidelines improves in the presence of a reference scan. The perceptual quality of the image improves with feedback. As a likely result, the accuracy of the manual measurements also improves. Accuracy of automated measurement remains consistently better than the manual measurements. Considerable effort is made to arrive at presets for specific clinical applications on commercial scanners to suit the needs of imaging representative patients for expert users. And yet, presets are not always optimal; the user often needs to alter either the acquisition or use knobs to alter the image for perceptual clarity. Through our experiments, we show that both accuracy and precision of automated measurements are not influenced by non-optimal presets. The results make a compelling case for enhanced automation and visual aids in addition to practice and feedback to lend higher reproducibility to ultrasonic measurements made by new users.

REFERENCES

- [1] Lopez, H., Loew, M.H. and Goodenough, D.J., "Objective analysis of ultrasound images by use of a computational observer," *IEEE Trans Med Imaging* 11(4), 496-506 (1992).
- [2] Rownd, J.J., Madsen, E.L., Zagzebski, J.A., Frank, G.R. and Dong, F., "Phantoms and automated systems for testing the resolution of ultrasound scanners," *Ultrasound in Med. & Biol.* 23(2), 245-260 (1997).
- [3] Kimberly, H.H., Murray, A., Mennicke, M., Liteplo, A., Lew, J., Bohan, J.S., Tyer-Viola, L., Ahn, R., Burke, T. and Noble, V.E., "Focused maternal ultrasound by midwives in rural Zambia," *Ultrasound in Med. & Biol.* 36(8), 1267-1272 (2010).
- [4] Mukherjee, P., Swamy, G., Gupta, M., Patil, U. and Krishnan, K.B., "Automated detection and measurement of femur length from fetal ultrasonography," *Proc. SPIE* 7629, 762909-762909-9 (2010).
- [5] Banerjee, J. and Krishnan, K., "Model-based detection of acoustically dense objects in ultrasound," *ICPR*, 4166-4169 (2010).
- [6] Annangi, P., Krishnan, K.B., Banerjee, J., Gupta, M. and Patil, U., "Automated detection and measurement of biparietal diameter from fetal ultrasonography," *Proc. SPIE* 7968, 796801-762901-9 (2011).
- [7] Liu, X., Annangi, P., Gupta, M., Yu, B., Padfield, D., Banerjee, J. and Krishnan, K., "Learning-based scan plane identification from fetal head ultrasound images," *Proc. SPIE* 8320, 83200A (2012).
- [8] Bryan, R.N., "The digital evolution, the millennium change in medical imaging," *Radiology* 229(2), 299-304 (2003).
- [9] Perni, S.C., Chervenak, F.A., Kalish, R.B., Magherini-Rothe, S., Predanic, M., Streltzoff, J. and Skupski, D.W., "Intraobserver and interobserver reproducibility of fetal biometry," *Ultrasound Obstet Gynecol* 24, 654-658 (2004).
- [10] Jardim, S. and Figueiredo, M., "Segmentation of fetal ultrasound images," *Ultrasound in Med. & Biol.* 31(2), 243-250 (2005).
- [11] Minitab Release 12.23 for Windows. Minitab Inc. PA 16801-3008, USA.