A case study in scanner optimisation

NJ Dudley¹ and NM Gibson²

¹Medical Physics Department, United Lincolnshire Hospitals NHS Trust, Lincoln, UK

²Medical Physics Department, Nottingham University Hospitals NHS Trust, Nottingham, UK

Corresponding author: Nick Dudley. Email: nick.dudley@ulh.nhs.uk

Abstract

Ultrasound scanner preset programmes are factory set or tailored to user requirements. Scanners may, therefore, have different settings for the same application, even on similar equipment in a single department. The aims of this study were: (1) to attempt to match the performance of two scanners, where one was preferred and (2) to assess differences between six scanners used for breast ultrasound within our organisation. The Nottingham Ultrasound Quality Assurance software was used to compare imaging performance. Images of a Gammex RMI 404GS test object were collected from six scanners, using default presets, factory presets and settings matched to a preferred scanner. Resolution, low contrast performance and high contrast performance were measured. The performance of two scanners was successfully matched, where one had been preferred. Default presets varied across the six scanners, three different presets being used. The most used preset differed in settings across the scanners, most notably in the use of different frequency modes. The factory preset was more consistent across the scanners, the main variation being in dynamic range (55–70 dB). Image comparisons showed significant differences, which were reduced or eliminated by adjustment of settings to match a reference scanner. It is possible to match scanner performance using the Nottingham Ultrasound Quality Assurance software as a verification tool. Ultrasound users should be aware that scanners may not behave in a similar fashion, even with apparently equivalent presets. It should be possible to harmonise presets by consensus amongst users.

Keywords: Ultrasound, optimisation

Ultrasound 2014; **22**: 21–25. DOI: 10.1177/1742271X13517231

Introduction

Modern ultrasound scanners usually have the facility to store settings appropriate to a particular probe or body part, so that they can be recalled by the user. The stored settings are usually recalled by selection of an appropriately named “preset,” e.g. Abdomen, Breast. Presets include a range of parameters, e.g. acoustic output, frequency, gain, dynamic range, image scale, focal depths, frame averaging, compound imaging and post-processing functions.

Ultrasound scanners are usually delivered with factory-defined preset programmes. Further presets may be defined locally to the requirements of individuals or groups of users, using a factory preset as a starting point. In our experience, it is not possible to include time gain compensation (TGC) settings in a locally defined preset. The ideal preset should be a good starting point for a particular type of examination. It might be reasonable to expect experienced ultrasound users to choose a similar starting point, so that there is a single preset per examination type and presets are similar across scanners.

In reality, scanners may be used with different settings for the same type of examination, even on equipment of the same model in a single department. Even where factory presets are chosen, these may be different on scanners of different ages as they are developed along with technological advances. Preset names are often reproduced across scanners but their content may vary widely.

On one site in our organisation, the three users expressed a preference for one scanner over another (similar model, different age); two users declined to use the older scanner. The aims of this study were: (1) to attempt to match the performance of the two scanners so that both gave similar images and (2) to assess differences between the six scanners used for breast ultrasound within our organisation. The Nottingham Ultrasound (US) Quality Assurance (QA) software¹ was used as a tool for comparison of imaging performance. It should be noted that studies investigating the relationships between test object results and the perception of clinical image quality have not demonstrated strong correlations,^{2,3} and it is reasonable to expect that the most recent models of a scanner will have improved clinical performance in comparison to older models.

Methods

Imaging performance was assessed using the Nottingham US QA system (version 8.2), measuring lateral resolution, low contrast (grey-scale) performance and high contrast (anechoic target detection) performance as follows: (1) lateral resolution, expressed as the full-width-half-maximum (FWHM) of the nylon filament images above the mean speckle background (mean speckle measured over 50% of target separation, typically about 20 pixels for a lower frequency probe in the ATS 539); (2) grey-scale target contrast, expressed as the ratio of mean pixel values on a circle inside and a semicircle proximal to the target image (0.7 and 1.35 times the radius of the target respectively); as image contrast increases, the slope of the plot of measured against specified contrast increases; (3) grey-scale target visibility, expressed as the standard error of the difference between the mean pixel values used for contrast measurement; (4) anechoic target visibility, using the correlation of an "ideal" image (black circle on white background) with the actual image. These measurements were chosen, as resolution and contrast are widely accepted as measures of imaging performance across all imaging modalities and, together with anechoic target visibility, are important to diagnostic efficacy in ultrasound.

The system also measures axial resolution, slice thickness and penetration depth. These measurements were not included in the study for the following reasons. Axial resolution is a small quantity, perhaps only two or three pixels in size and hence has a large variation due to measurement uncertainties. Slice thickness is largely probe-dependent and is expected to be fixed for probes of the same model, although it should differ between fundamental and harmonic imaging. The clinical settings used meant that penetration depth was beyond the imaged depth (confirmed by image analysis using increased range setting).

Images of a Gammex 404GS test object (Gammex RMI, Middleton, WI, USA) were collected (in DICOM format on CD) using the default clinical settings for each scanner used in the Breast Screening Units of our organisation. Table 1 shows details of the scanners (Toshiba Medical Systems Corporation, Tochigi-ken, Japan) and their original default presets. Each scanner used a PLT-1204AT linear array probe. For all images in the study TGC slider controls were set to their central position.

Table 1 Ultrasound scanners in the study

Scanner code	Model	Year of purchase	Default preset
L1	Toshiba Aplio SSA790A	2007	Breast 1204
L2	Toshiba Aplio SSA770A	2004	Breast Harmonic
B1	Toshiba Aplio SSA790A	2008	Breast 1204
B2	Toshiba Aplio SSA790A	2007	Breast (Factory)
G1	Toshiba Aplio SSA780A	2011	Breast (Factory)
G2	Toshiba Aplio SSA770A	2009*	Breast 1204

*G2 manufactured in 2006.

An initial comparison was made between scanners L1 and L2, where scanner L1 was subjectively preferred over scanner L2 by the users. The original settings are shown in Table 2. Users were unable to detail the reasons for this preference in objective terms, but the difference in dynamic range shown in Table 2 should have an effect on contrast. A preset was then created on scanner L2 to match the default settings on scanner L1, and the comparison repeated. For each scanner and preset, five sets of images were collected and the mean results compared.

Comparisons were then broadened to include all six scanners in the Breast Screening Units, images from each scanner being obtained with each of the two presets most commonly used. Comparisons were made to assess whether the presets contained different settings on scanners of different ages and on different sites and whether measured image parameters were consistent between scanners using the same preset or matched settings. For image comparisons, L1 was used as a reference and a single set of images for each of the other scanners compared with this reference (mean of five measurements ± 2 standard deviations).

Background grey-scale levels at each target depth were available when measuring resolution; these are not routinely used, but were recorded to allow any necessary further assessment of differences between images.

Results

The measured imaging performance of scanners L1 and L2 differed initially and matching the preset of the two scanners resulted in equivalent performance in a test object. Figure 1 shows results for lateral resolution, where there is a significant difference ($p < 0.05$) up to a depth of 30 mm between L1 and L2 using the original default settings and no significant difference ($p > 0.05$) after adjustment. Low and high contrast results were similar, with significant differences before adjustment and no significant differences after adjustment ($p < 0.05$).

Assuming that scanner L1 was set for optimum clinical performance, we had therefore optimised scanner L2 (with the caveat that a test object may not truly represent the clinical situation). All three users were subjectively more satisfied with the performance of scanner L2 than previously and began to use it again.

Table 2 shows key settings for each of the six scanners when using their version of the Breast 1204 preset. Range defaulted to 40 mm and output to maximum in each preset; for our initial measurements gain was adjusted for a speckle level of approximately 30% of peak white irrespective of the default. Automatic gain control (AGC) defaulted to off and edge enhancement (setting 1) and 2D map (grey-scale transfer curve; setting 2) were at the same level on all scanners. No comparison between images at these settings was made as there were several differences in settings. Most notably, different default frequencies were used, being set according to user preference on installation.

Comparisons between images for the six scanners when adjusted to use the Breast 1204 settings found on L1 (including overall gain) showed only two significant

Table 2 Default settings for the Breast 1204 preset for each of the scanners (L2 later set to match L1; G1 has Precision instead of post processing, setting 2). Aplipure settings include spatial and frequency compounding

	L1	L2	B1	B2	G1	G2
Frequency (MHz)	T14	T14	Diff13	Diff13	Diff13	14
Dynamic Range (dB)	60	70	70	70	65	60
Focal Depths (mm)	10, 20	15, 20	7.5, 15	10, 20	15	7.5, 15
Aplipure	1	0	0	0	4	N/A
Post Processing	6	6	6	6	N/A	2
Time Smooth	3	4	5	5	2	3

T: tissue harmonic mode; Diff: differential harmonic mode; N/A: not available.

Figure 1 Lateral resolution for scanner L1 (Breast 1204), L2 with original default preset (Default; Breast Harmonic) and L2 with preset adjusted to that of L1 (Adjusted; Breast 1204). Mean of five measurements shown; error bars show ± 1 standard error (SE)

differences in performance. Measured contrast for G2 showed a slightly reduced slope relative to L1. The visibility of the 2 mm anechoic target was slightly less than the reference range for B1; the other targets were within the reference range and so this may be an artefactual result. In summary, most results matched the reference scanner after adjustment to harmonise the settings.

A notable finding when making the adjustments for these measurements is that the scanners behave differently on changing mode, e.g. adjusting from differential harmonic imaging to standard harmonic imaging on B1 and B2 resulted in different dynamic ranges and focal numbers and depths; switching from harmonic to fundamental imaging on L1 and L2 resulted in very different dynamic ranges (55 dB and 70 dB respectively) and different focal numbers and depths.

Table 3 shows key settings for each of the 6 scanners using their version of the Breast Factory preset. Range defaulted to 40 mm and output to 20 in each preset; for our measurements output was adjusted to maximum and

Table 3 Default settings for the Breast Factory preset for each of the scanners (G1 has Precision instead of post processing, setting 0)

	L1	L2	B1	B2	G1	G2
Frequency (MHz)	14	14	14	14	14	14
Dynamic range (dB)	60	70	55	60	65	65
Focal depths (mm)	20	20	20	20	20	20
Aplipure	0	0	2	0	0	N/A
Post processing	2	2	2	2	N/A	2
Time smooth	5	4	5	4	5	4

N/A: not available.

gain adjusted for a speckle level of approximately 30% of peak white irrespective of the default. Automatic gain control (AGC) defaulted to off and edge enhancement (setting 1) and 2D map (setting 5) were at the same level on all scanners. The most notable difference was in dynamic range.

Comparisons between images for the six scanners using the Breast Factory preset showed no significant differences

Figure 2 Low contrast (grey-scale) performance for Breast Factory preset adjusted to L1 settings. $\pm 2SD$ range shown for reference scanner (L1). Results from single images shown for other scanners. The target with greatest negative contrast is scatter-free but labelled here as -20 dB for convenience

in resolution, but significant differences in low contrast (grey-scale) and high contrast (anechoic target detection) performances. Adjusting the settings to match L1 (including overall gain) reduced the number and level of significance of results differing from L1, so that only low contrast performance showed significant differences (Figure 2).

The additional background grey-scale level results obtained with resolution measurements showed that in the well-matched performance measurements (the initial comparison of L1 and adjusted L2 and the comparison of scanners with Breast 1204 preset adjusted to match L1) there were no significant differences ($p < 0.05$) between mean background grey levels between the sets of images. In the images taken using the Factory preset adjusted to match L1, Figure 3 shows that there were significant differences in background grey-levels reflecting a difference in sensitivity between the scanners at these settings. The most significant difference in low contrast performance results was for B2; the grey level at the depth of the targets (35 mm) was 23% higher for B2 than for L1. Figure 3 also shows that the grey-scale gradient across targets varies between scanners and this will affect contrast results (background signal is measured superficial to the targets on a semicircle from approximately 30 mm to 35 mm depth and the targets are centred at approximately 35 mm).

Discussion

Presets on what appear to be similar scanners may not be the same and adjustments of parameters such as frequency mode may have different effects. Differences between presets may be, for good reasons, related to software or hardware enhancements between models/versions. However,

differences in the effects of adjustments may lead to problems with image quality unless the user is aware and can compensate; there could be implications if staff move between centres and expect all scanners to behave in the same way.

In the comparison between L1 and L2, results were closely matched following adjustment of L2. Resolution appears to have deteriorated after “optimisation”; the reduction in dynamic range setting from 70 dB to 60 dB altered the filament profile (increasing full-width-half-maximum). Test object results are not necessarily representative of clinical performance, and dynamic range should be set to suit the clinical task.

Breast 1204 settings are locality specific, i.e. they have been adjusted to the preference of local staff in discussion with the supplier’s clinical applications specialist. Adjusting them to match a reference scanner (L1) gave similar results in test object measurements.

Factory presets change as new technologies are introduced or modified, so the differences in results between the scanners were not unexpected. We achieved closer, but not identical, performance by adjusting factory settings to a reference. Our success in matching performance on the Breast 1204 preset but not on the Factory preset was due to differences in grey-scale levels using the latter preset.

The significant differences in background grey levels shown in Figure 3 show that there are settings other than those we have adjusted that affect grey levels. Differences in grey levels varied with depth, so it may be that preset TGC differs between scanners.

A harmonised approach should be possible. Sonographers and radiologists should agree a common starting point for a particular type of examination and

Figure 3 Mean background grey levels for each scanner after adjustment to the reference Breast Factory preset (L1; mean $\pm$ 2SD). Target depth was approximately 35 mm

this should be programmed into a preset. Some parameters should be straightforward and transferable between scanners, such as frequency, whether harmonics should be used from the outset, range and focal depths. The setting of other parameters, such as dynamic range, may be more dependent on the age and model of the scanner and some may not be reprogrammable, e.g. TGC.

The scanners in this study are used in the United Kingdom National Health Service (NHS) Breast Screening Programme. If a locally harmonised approach is possible, then screening programme leads should consider national harmonisation. To the authors' knowledge this has been attempted in the NHS only in the Abdominal Aortic Aneurysm Screening Programme, where the use of tissue harmonic and compound imaging is recommended.⁴

Conclusions

We have been able to match imaging performance (in a test object) by harmonising scanner settings. The Nottingham US QA system provided a useful verification tool in comparing performance between similar scanners from the same manufacturer. Matching scanners of different manufacturers or perhaps even significantly different models is a more complex process and requires work to assess feasibility. However, matching scanner performance is only true optimisation if the reference scanner has been set to give the optimum clinical performance, which is currently a subjective process.

The next stage of this study will be to work with ultrasound practitioners to try to understand their requirements and move towards harmonisation of presets.

DECLARATIONS

Conflict of interests: The authors have no conflicts of interest to declare.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical approval: Not required

Guarantor: NJD

Contributorship: NJD researched and conceived the study and performed the measurements. NMG developed the Nottingham US QA software. NJD wrote the first draft of the manuscript. Both authors reviewed and edited the manuscript and approved the final version.

REFERENCES

1. Gibson NM, Dudley NJ, Griffith K. A computerised ultrasound quality control testing system. *Ultrasound Med Biol* 2001;**27**:1697-711
2. Metcalfe SC, Evans JA. A study of the relationship between routine ultrasound quality assurance parameters and subjective operator image assessment. *Br J Radiol* 1992;**65**:570-5
3. Browne JE, Watson AJ, Muir C, et al. An investigation of the relationship between in vitro and in vivo ultrasound image quality parameters. *Ultrasound* 2004;**12**:202-10
4. Essential elements in developing an abdominal aortic aneurysm (AAA) screening and surveillance programme. NHS Abdominal Aortic Aneurysm Screening Programme, 2011