The Resolution Integral as a metric of performance for diagnostic grey-scale imaging

S D Pye and W Ellis
Medical Physics Department, NHS Lothian University Hospitals Division, Royal Infirmary, Edinburgh EH16 4SA, UK
E-Mail: stephen.pye@luht.scot.nhs.uk

Abstract. There is a strong clinical need for objective measurements of grey-scale ultrasound imaging performance. We have previously proposed the Resolution Integral \( R \) as a suitable metric, because it defines the ratio of the penetration of an ultrasound beam in soft tissue to the ultrasound beam width. Transducers with good performance combine deep penetration with high lateral resolution, giving a large ratio of penetration to beam-width. Depth of Field \( L_R \) and Characteristic Resolution \( D_R \) are defined such that \( R = \frac{L_R}{D_R} \). \( L_R \) defines a region of optimum resolution, and \( D_R \) is representative of the lateral and elevation resolution within \( L_R \). We report the results of a survey of the imaging performance of 79 models of ultrasound scanner made using an Edinburgh Pipe Phantom to measure values of \( R \), \( L_R \) and \( D_R \). The scanners were manufactured between 1986 and 2009 and were from 23 manufacturers. A total of 171 different transducers were tested, including linear and convex arrays, and mechanical sector probes. The results demonstrate that \( R \) successfully distinguishes transducers with differing levels of performance, and that \( L_R \) and \( D_R \) characterise probes suitable for different clinical applications. The characteristics of individual probes are concisely quantified and displayed on a plot of \( L_R \) vs \( D_R \).

1. Introduction
The objective assessment of diagnostic ultrasound imaging performance has several roles: to ensure that equipment performs to a defined standard and is clinically fit-for-purpose; to assess new imaging modalities and signal processing techniques; to underpin decision making in the procurement and replacement of equipment. Technical assessments of grey-scale ultrasound have been limited by the fact that the resolution of ultrasound images is a strong function of depth and varies in three orthogonal directions [1]. As a consequence, measurements of spatial and contrast resolution made with test objects often produce large volumes of data which may be useful to monitor equipment during its life cycle [2], but are restricted in their ability to compare the performance of different ultrasound systems. International standards are limited to describing procedures for testing the accuracy of measurement callipers and for measuring resolution and point spread function at discrete depths [3, 4]. The limitations of current methods for performance evaluation are discussed by Shaw and Hekkenberg [5].

Pye and Ellis [6] introduced the concept of a Resolution Integral and published preliminary work [7]. This was the first description of a fundamental physical approach to combining measurements of resolution made at different depths in the ultrasound image. Moran et al [8,11] describe the use of the Resolution Integral to characterise pre-clinical ultrasound scanners, and Rowland et al [9] used a test...
phantom containing a step-change in backscatter to make independent measurements of Resolution Integral in the scan and elevation planes. In this paper we report the results of a survey of ultrasound equipment carried out in Lothian Region, Scotland between 2002 and 2010. Measurements of Resolution Integral were made using a number of Edinburgh Pipe Phantoms [6, 7]. The aim of the survey was to determine whether measurements on individual transducers could characterise their performance and suitability for their specified clinical application. In particular, whether R could discriminate improvements in transducer technology, and whether $L_R$ and $D_R$ could characterise clinical application.

2. Resolution Integral

The Resolution Integral $R$ for a collimated beam is defined as the ratio of penetration to beam width [7]. For real ultrasound beams, the variation of beam characteristics with depth is accounted for using an integral approach. Figure 1a illustrates this concept. The definition of $R$ has been amended slightly here, so that it is defined in terms of beam width rather than beam radius as in [7]. Two other parameters are defined in figure 1b: Depth of Field $L_R$ and Characteristic Resolution $D_R$. The Depth of Field defines the axial extent of a region of optimum imaging, analogous to the focal region. The Characteristic Resolution is representative of the beam width within the Depth of Field. Medical ultrasound transducers with good performance combine deep penetration with high resolution, so that the ratio of penetration to beam width is large. A transducer suited to a particular clinical application will have a small Characteristic Resolution together with sufficient Depth of Field for the particular application.

![Figure 1a](image1a.png) **Figure 1a.** The graph (left) shows a plot of $L$ vs $\alpha$ for an ultrasound beam (right) with low contrast penetration $L_0$ and minimum beam width $D_0$. $\alpha$ is the reciprocal of beam width, and $L(\alpha')$ corresponds to the depth range in the image over which the beam width is less than $1/\alpha'$. The Resolution Integral $R$ is equal to the area under the curve.

![Figure 1b](image1b.png) **Figure 1b.** The graph (left) shows the same plot of $L$ vs $\alpha$ as Fig 1a, and illustrates Depth of Field $L_R$ and Characteristic Resolution $D_R$. The area under the dashed curve $0-L_0-1/D_0$ is equal to the Resolution Integral $R$. The rectangle $0-L_R-0'-1/D_R$ also has an area equal to $R$, and line $00'$ bisects both the rectangle and the area under the dashed curve, giving $R = L_R / D_R$

3. Materials and Method

Measurements were carried out on 79 different models of ultrasound scanner manufactured between 1986 and 2009, and sited at various hospitals and clinics in Lothian Region. These included top- and mid-range scanners and a smaller number of portable and hand-held machines, employing linear arrays, convex arrays and mechanical sector probes. A total of 171 different scanner/transducer combinations from 23 manufacturers were assessed: 78 abdominal transducers, 57 small parts transducers, 23 transvaginal (TV) endoprobes, 9 cardiac transducers, 3 vascular access imagers, and 1...
th bladder imager. Scanners were either in clinical service when tested, withdrawn from service and maintained for laboratory use, or undergoing pre-purchase evaluation. They were all in good working order with no known faults.

The measurements of $R$ were made by three operators using six Edinburgh Pipe Phantoms, manufactured in-house as described in [7]. Each pipe phantom was constructed from a 250 x 250 x 100 mm block of agar based tissue mimicking material (TMM), prepared using the method described in [10]. The TMM had an attenuation coefficient of 0.5dB cm$^{-1}$ MHz$^{-1}$ and speed of sound of 1540ms$^{-1}$. Eleven wall-less pipes ranging in diameter from 0.35mm to 7.9mm were moulded into the TMM at an angle of 40° to the vertical and filled with water-glycerol mixture ($c=1540$ms$^{-1}$) [7]. The six phantoms were cross-compared using small parts and abdominal probes. Measurements of $R$ made using the same probes on different pipe phantoms were within a 5% range. The stability of the pipe phantoms over time was tested by making periodic measurements of $R$ during the course of the survey using the same scanner/transducer combination. Results were within the repeatability of the measurement (3%).

Scanning took place under a controlled level of ambient lighting (<20 lux at the surface of the display monitor). Each transducer was placed on the surface of the pipe phantom and coupled to it with water/glycerol mixture. An image of one of the pipes was obtained, and the upper portion of the pipe was imaged with the pipe axis positioned in the scan plane of the transducer. The image was optimised to visualise the pipe as superficially as possible, with the transducer held vertically, and the most superficial region of the pipe that could be detected positioned directly below the transducer (within ±10° of the vertical). All available controls were used to optimised the image e.g. field of view, TGC, transmit focus, transmit frequency, harmonic imaging, speckle reduction, compound imaging. The image was then frozen and a visual assessment was made of the minimum depth at which the pipe could be visualised. To aid this assessment, a mask of buff-coloured paper with a slot of width approximately 15 times the wavelength of the centre frequency of the probe was used to mask out echoes from neighbouring regions of the pipe and TMM. The mask was placed onto the frozen ultrasound image and the slot was moved up and down along the length of the pipe image to allow the observer to visualise each short section of the pipe in turn, compare it with adjacent speckle, and identify and measure the minimum depth below the scanning surface at which the pipe could confidently be detected. The transducer was then re-positioned, the image was optimised, and the same procedure was carried out to identify the maximum depth below the scanning surface at which the pipe could confidently be detected. The difference between the maximum and minimum depths at which the pipe could be detected gave one value of $L$ (figure 1). This procedure was repeated for each pipe and each value of $L$ was plotted against $\alpha$, the reciprocal of the effective pipe diameter. The effective pipe diameter was calculated as the geometric mean of the pipe diameter in the image and elevation planes and is equal to $d/\cos(40^\circ)$ for a pipe of diameter $d$. The low contrast penetration (LCP) of each transducer was determined by imaging in real time and measuring from the scanning surface to the deepest speckle identifiable from noise. Care was taken to ensure that no pipes were visible in the field of view. The depth of the transducer dead zone (DZ) was measured vertically below the probe. The distance LCP – DZ represents the maximum range of depth over which anechoic objects of any size can be detected, and corresponds to the intercept ($L_\alpha$) on the ordinate in figure 1. Typically, the measured $L-\alpha$ plot for each transducer comprised values of $L$ from 6 - 7 pipes plus the distance LCP – DZ. The area under the $L-\alpha$ curve was integrated using a simple trapezium rule to calculate $R$. Values of $L_R$ and $D_R$ as defined in figure 1b were also calculated.

4. Results

Figure 2 shows the measured values of $R$ for each transducer/scanner combination, plotted against year of manufacture and grouped in 5-year intervals. The median value of $R$ is shown for each group, together with the interquartile range. The median value for 2005-09 is significantly different from 2000-04 ($P<0.04$); as is 05-09 vs 95-99 ($P<0.01$); and 00-04 vs 90-94 ($P<0.01$). As an illustration of some of the data summarised in figure 2, figure 3 shows values of $R$ measured for a series of scanners.
Figure 2. Median value of Resolution Integral and interquartile range, plotted against year of manufacture for 171 transducer/scanner combinations grouped in intervals of five years. N denotes the number of transducers in each group.

Figure 3. Resolution Integral values for different models of scanner manufactured between 1994 & 2008 by Toshiba Medical (left) and Hitachi Medical (right). Older scanners are on the left of each chart and the newer ones are on the right. Each bar represents a single transducer. Solid bars = abdominal probes. Hatched bars = TV endoprobes. 95% confidence intervals are shown.

Figure 4. Plot of $L$ normalised by $L_R$ versus $\alpha$ normalised by $D_R$ for 171 transducer/scanner combinations. All the curves lie close to each other, indicating that $L_R$ and $D_R$ are efficient parameters for describing beam shape. i.e. Probes with the same values of $L_R$ and $D_R$ have very similar beam shapes; probes with different values of $L_R$ and $D_R$ have different beam shapes.
produced over a period of 15 years by two major manufacturers. To test the validity of using $L_R$ and $D_R$ to characterise beam shape, the measured values of $L$ for each transducer were divided by $L_R$ for that transducer, and the values of $\alpha$ were divided by $1/D_R$ to produce a normalised plot of $L/L_R$ against $\alpha D_R$. The results for all 171 transducers are shown together in figure 4.

Figure 5 shows a schematic plot of $L_R$ vs $D_R$, and figure 6 illustrates the results obtained for 158 probes designed for use in abdominal imaging, small parts imaging, and transvaginal scanning.

**Figure 5.** Schematic plot of $L_R$ versus $D_R$. Two transducers are shown: a small parts probe located in the lower left quadrant, and an abdominal probe in the upper right quadrant. The value of $R$ for each probe is equal to the gradient of the line joining it to the origin ($R = L_R/D_R$). Equipment with low values of $R$ (“old technology”) is located in the lower right quadrant. Equipment with high values of $R$ (“new technology”) is located in the upper left quadrant.

**Figure 6.** Measured values of $L_R$ and $D_R$ for 158 transducers designed for three different clinical applications. The dashed lines represent lines of constant $R$ (25, 50 and 75). Modern premium range equipment lies close to or above the line $R=75$. The oldest transducers and those with the most limited performance lie close to $R=25$.

5. **Discussion**

The Resolution Integral combines a number of ultrasound image parameters (resolution at different depths, low contrast penetration, dead zone) in a physically meaningful way. The results shown in figure 2 demonstrate that $R$ is able to characterise improvements in ultrasound imaging technology that have taken place over the last 25 years. This is supported by the data in figure 3, where the values of $R$ for a series of scanners produced by two manufacturers increase between older models and modern high-end equipment (from 40-50 to >80 for abdominal probes; from 30-40 to 50-60 for TV endoprosbes). One limitation of the survey is the uncertainty regarding any historical deterioration in performance, given that some scanners were >10 yrs old when tested. The majority of older scanners
were part of a local quality assurance programme and had been assessed at regular intervals using commercial tissue mimicking test objects. Although this data did not provide useful information about resolution, measurements of low contrast penetration had been consistent over time since installation. Given that the equipment was maintained in good working order, we estimate that any change in the Resolution Integral over the lifetime of the scanners was $<10\%$. It is also useful to compare the Resolution Integral values of five single-element transducers tested during the survey. Two non-diagnostic imagers had $R$ values of 22 and 24 (Bardscan bladder imager, manufactured 2005; Dymax SiteRite II, manufactured 1994). Three single element diagnostic probes had $R$ values of 36, 36.7 and 36.8 (Siemens Sonoline AC CDA3.5, 1991; Siemens Sonoline SX CDA3.5, 1986; Kretz Combison 401, 2002). The different $R$ values of the diagnostic and non-diagnostic equipment suggest a difference in technology rather than a historical deterioration in performance.

Figure 6 shows $L_R$ and $D_R$ values for transducers designed for abdominal, transvaginal and small parts imaging. Transducers designed for each clinical application are grouped together in different parts of the graph, as would be expected by comparison with figure 5. Three abdominal probes shown in figure 6 with $D_R \sim 1.1$mm appear to be outliers, but these were in fact paediatric transducers operating at a higher centre frequency than probes designed for adult work. The median values of $R$ for the three groups of transducers in figure 6 are: 62 (abdominal), 49 (TV endoprobes), 58 (small parts). The low value of $R$ for TV endoprobes may be due to the small radius of curvature of the imaging arrays used on many of the probes. This limits the size of the acoustic aperture, and also generates rapidly diverging scan lines. These two factors may restrict the improvements that can be achieved in image quality. This hypothesis is supported by the data in figure 3, which shows a smaller increase in $R$ for older & newer TV endoprobes, compared to older & newer abdominal transducers.

6. Conclusions
The Resolution Integral, measured using an Edinburgh Pipe Phantom, is able to discriminate between medical ultrasound transducers with differing levels of performance. The Depth of Field $L_R$ and Characteristic Resolution $D_R$ provide a quantitative summary of the imaging characteristics of individual probes that can be concisely displayed on a plot of $L_R$ vs $D_R$.

Acknowledgements
The authors wish to thank Anna Janeczko, Carmel Moran, George Campbell and Scott Inglis for their assistance.

References
[1] Madson E L 2000 *Ultrasound in Med. & Biol.* 26 Supplement 1 ppS48-S50
[2] Thijsen J M, Weijers G and de Korte C L 2007 *Ultrasound in Med. & Biol.* 33 460-71
[3] International Electrotechnical Commission 1996 IEC 1390
[4] International Electrotechnical Commission 2006 IEC 61391-1
[5] Shaw A and Hekkenberg R 2007 NPL Report AC 2. UK National Physical Laboratory, London
[6] Pye S D and Ellis W 2002 UK Patent GB2396213
[7] Pye S D, Ellis W and MacGillivray 2004 *Journal of Physics: Conf. Series* 1 187-92
[8] Moran C M, Smart S, Ellis W and Pye S D 2008 *Proc 2008 IEEE Ultrasonics Symposium* pp1724-27
[9] Rowland D E, Newey V R, Turner D P and Nassiri D K 2009 *Ultrasound in Med. & Biol.* 35 1525-34
[10] Inglis S, Rammarine K V, Plevris J N and McDicken W N 2006 *Ultrasound in Med. & Biol.* 32 249-59
[11] Moran C M, Ellis W, Janeczko A, Bell D and Pye S D 2010 *Journal of Physics: Conf. Series* (this issue)