THE UNIVERSITY *of* EDINBURGH
School of Physics
and Astronomy

Investigating the Effect of Slice Thickness on Ultrasound Imaging Performance

MPhys Project Report

David Roddy

Submitted for the 20pt MPhys Project with Year Abroad course PHYS11050
March 27, 2023

Abstract

The effect of slice thickness on ultrasound image quality is not well quantified. The resolution integral is an established technique for calculating a single figure-of-merit that combines imaging performance over an entire ultrasound beam. Using a commercial slice thickness phantom, continuous slice thickness versus depth profiles were recorded for 10 transducers and the resolution integral was adapted to quantify imaging performance solely in the elevation plane. Slice thickness was identified as having a significant impact on ultrasound imaging performance but no clear relationship was found between the calculated resolution integrals, likely due to confounding factors from the lateral plane. There was a strong linear relationship between the sizes of the overall and elevational focal regions, while the typical resolutions were correlated but also affected by confounding factors.

Supervisors: Professor Carmel Moran, Dr Scott Ingles, Dr Stephen Pye

Personal statement

I started this project by reading about the resolution integral and Edinburgh Pipe Phantom (EPP). My first two weeks involved being taught how to operate an ultrasound scanner and testing out the different settings. This involved imaging test phantoms used for science outreach, and following checklists to make sure I understood how each setting affected the image. I then began reading about the slice thickness phantom and filling in risk assessments for working at the hospital. I also began working on some code to calculate the resolution integral from EPP measurements.

In week 4 I used the EPP for the first time and took some measurements with it. I repeated this for a few probes, getting familiar with the EPP measurement process. I used the code to compare my measurements to those of the experienced operators. By week 7 we were putting together a plan for taking videos with the slice thickness phantom using a rig for smooth movement.

This plan was then executed and I had 12 videos with two different probes to work with. I wrote code to analyse the videos and return a slice thickness profile. At this point, we wanted to combine this data with EPP results but were not sure exactly how. I had an idea that maybe we could use the slice thickness profiles to calculate a resolution integral encompassing only the slice thickness and I did some preliminary analysis to see if that could be done. I discussed this with my supervisors and we decided to move ahead with it after working out some of the details about the scanner settings. Just before the holidays, I took videos with two more transducers.

Over the break, I worked on the analysis code and analysed all the videos so far. I had not automated many of the steps of the code so analysis took a long time at this point. I obtained resolution integral values for the first 4 transducers and did some preliminary comparisons before taking videos with four more probes. I then found a mistake in my analysis and had to re-analyse all the data so far. I then took videos with several more probes, including repeats where I had used different scanner settings.

I automated the analysis since I then had nearly 70 videos. I took some videos with two probes using a different phantom to compare, but the data was too noisy and so was omitted from this report.

I began writing this report and making further improvements to the code, before re-analysing all videos with the latest code. The last few weeks focused on writing the report, estimating uncertainties, and fully validating the resolution integral code against the spreadsheet.

Acknowledgments

I would like to thank Carmel Moran, Scott Inglis, and Stephen Pye for their support and guidance throughout this project. Thank you also to Christopher McLeod and the Medical Physics department at the Royal Infirmary Edinburgh for their patience and help.

Contents

1 Introduction

Ultrasound is one of the most widely-used imaging modalities, with more ultrasounds conducted in 2016 in the United States than computed tomography scans, magnetic resonance imaging or nuclear medicine imaging, and the number of scans is increasing year on year [1]. With this intense level of usage, it is extremely important to monitor the imaging performance of ultrasound systems to keep patients safe. Clinical staff also assess imaging performance to improve decision-making when new equipment is purchased and to compare new technology against the current state-of-the-art. For these reasons, there are many methods and test objects which offer ways to quantify ultrasound performance, with many hospitals implementing their own procedures [2, 3]. However, ultrasound resolution is different depending on the depth of imaging, so evaluating imaging performance can be complicated to achieve and interpret. The resolution integral was introduced in 2004 and became the first technique to produce simple, physically relevant results which accounted for variation with depth [4]. The resolution integral has since been successfully used for quality assurance and research [5–10]. Ultrasound beams are three-dimensional, and the resolution integral combines them all when calculated using the Edinburgh Pipe Phantom (EPP) which makes it a useful measure of overall imaging performance. It is known that the dimension perpendicular to the scan plane (the elevational dimension or slice thickness) affects the imaging performance and therefore the resolution integral, but the exact nature of this effect is not fully understood. This report details an attempt to discover more about the relations between overall imaging performance and variation in the slice thickness (ST) plane by calculating a resolution integral for the ST plane only. In the background, an overview of medical ultrasound and beam characteristics are given before more details about the resolution integral and the known effects of ST. The procedure for gathering data, analysing it, and calculating the new resolution integrals is then explained. Several results obtained at various stages of the method are given along with the full calculated resolution integrals and comparisons to EPP data. These results are then interpreted and put into context before conclusions and future outlooks are discussed.

2 Background

2.1 Medical ultrasound

Medical ultrasound is a non-invasive diagnostic imaging technique that uses sound waves above the frequencies of human hearing to create an image of internal organs and tissues. Ultrasound is a safe and painless procedure that does not use radiation, making it ideal for people who may not be suitable for radioimaging such as pregnant women and children. It is also a relatively quick and inexpensive imaging method that can provide valuable diagnostic information to clinical staff in real-time, allowing them to make faster and more accurate diagnoses. The use of ultrasound in a medical setting arose from underwater sonar research after the Second World War, and nowadays it is widely used for purposes including assessing

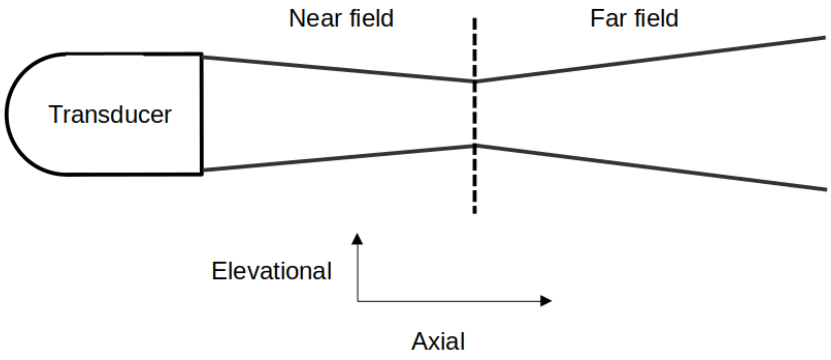


Figure 1. Ultrasound beam shape in the elevation plane. Note the convergence to the focal depth followed by indefinite divergence.

the health of developing fetuses during pregnancy, diagnosing and monitoring conditions affecting the heart and blood vessels, and identifying abnormalities or injuries in organs such as the liver, kidneys, and brain [11].

There are two main components of a medical ultrasound system: the scanner and the transducer. Ultrasound transducers contain at least one piezoelectric element (typically a ceramic), which converts electrical energy into mechanical energy (sound) through the deformation of its structure. Conversely, piezoelectric crystals also produce a voltage difference when they are deformed, allowing them to both transmit and receive ultrasound waves. The waves travel into the tissue, getting scattered, reflected, refracted, and absorbed. Signals which return to the transducer are interpreted and displayed by the scanner. While older transducers had simple elements which could only transmit at a resonant frequency defined by the size of the element, modern transducers contain specially machined elements which enable transmission at a variety of frequencies besides this 'central frequency'. Most ultrasound systems have transducers with between 64 and 512 elements arranged in a flat row (a linear array), several flat rows (a multi-row or 1.25D array), or a curved row (a curvilinear array). To build up an image, the pattern in which these elements are activated can be either "linear" or "phased". Phased arrays activate all elements nearly simultaneously and steer the beam using slight delays in activation for particular elements. Linear arrays instead activate small groups of around 20 elements in a sweep along the row. The image is created by combining the returning signals from each of the beams, it will be rectangular for linear arrays and arced-trapezoidal shaped for curvilinear arrays.

2.2 Ultrasound beam shape and characteristics

The shape of an ultrasound beam emitted by a transducer is shown in Figure 1. Key features are the convergence of the beam in the near field and divergence in the far field. The depth at which the beam begins to diverge (the focal depth) is determined by the size of the elements and the frequency of waves they are emitting. The convergence in the near field is caused by

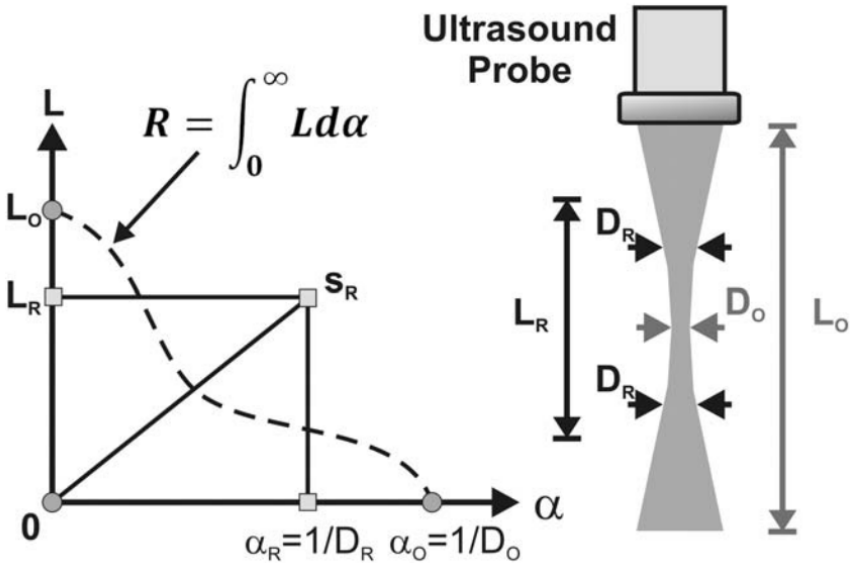


Figure 2. A geometric demonstration of the resolution integral. The graph on the left shows the visualisation depth (L) against the inverse diameter of the imaged pipe (α), also known as an L-α plot. The schematic on the right shows how the quantities marked on the plot match up to the physical characteristics of the ultrasound beam. Reproduced from [14]

interference patterns, and the beam's width is approximately half that of the transducer's aperture diameter at the focal depth [11]. The point of peak ultrasound pressure is at the focal depth, as is the best resolution - and therefore the clearest image. While the intrinsic focal depth is not changeable, most transducers can adjust the timing of element activation to focus the beam at different depths, or even at multiple depths simultaneously. The multi-row arrays mentioned earlier can also focus their beams in the elevation plane, improving the ST resolution and enabling an even clearer image. Image quality is further improved by receive focusing, where detected signals are selected by time of arrival to focus at a specific depth. Using multiple focus depths comes with a frame rate penalty, so a balance must be found if real-time smooth imaging is required.

The ultrasound frequency strongly affects the beam characteristics. Higher frequencies have a shorter wavelength which improves the axial resolution and image detail, but the signal is more strongly attenuated leading to a significantly reduced penetration depth [12]. This means higher frequencies (>7.5 MHz [13]) are generally used for small parts imaging for organs close to the surface, while lower frequencies are typically used for purposes requiring greater penetration depth such as abdominal imaging. Modern transducers often have the ability to detect higher harmonics of the original frequency which can improve the detail of the image.

2.3 The resolution integral

The imaging performance of ultrasound scanners is important for maintaining a high standard of healthcare. Ensuring that damaged scanners are detected and repaired, evaluating new technologies, and choosing the best system for a purpose all depend on measurements of the quality of ultrasound imaging performance. Some important aspects to be monitored are the dead zone, axial resolution, lateral resolution, and ST resolution [3, 13, 15].

The first proposed method to combine resolution measurements made at various depths was the resolution integral (R), which produces a single figure allowing quick comparisons between different ultrasound systems [4]. This technique has since been used to evaluate performance changes over time and also between types of ultrasound scanner [5–10]. The resolution integral can be calculated by making use of the Edinburgh Pipe Phantom (EPP) which contains around eight anechoic pipes of varying size [14]. Measurements are made of the depth range (L) over which a pipe of a given diameter can be resolved, which allows a curve of L against α (1/pipe diameter) to be drawn as in Figure 2. The resolution integral can then be found with $\mathrm{R} = \int_0^\infty \mathrm{L}\ d\alpha$ by interpolating between measurements. Two further figures of interest can be derived from this, the characteristic resolution ($\mathrm{D_R}$) and the depth of field ($\mathrm{L_R}$). A rectangle is constructed with the same area as the area under the L-α curve and a diagonal which bisects the area under the curve. The extent of this rectangle in the L and α dimensions represent an ideal beam with the same R as the real beam, and these values are termed the characteristic resolution and the depth of field. They are analogous to the extent of the focal region and the typical resolution within it.

2.4 Slice thickness

The ST of an ultrasound beam is known to affect the resolution integral, but this is yet to be fully quantified. In a previous study, a modified version of the EPP using metal bars was used to obtain R for several transducers and found that R increased by a factor of at least 2.5 when the effect of ST was removed [16]. However, there was significant uncertainty (up to 30%) in the exact value of R, and by removing the effects of ST it was not possible to explore the details of the relationship between ST and overall imaging performance. This study attempts to do almost the opposite and isolate ST to gain a better understanding of exactly how it influences imaging performance.

Test objects have been developed to measure ultrasound beam ST specifically which contain a scattering plane at 45° to the top and bottom surfaces [17, 18]. Coupling a transducer to the surface of the phantom shows a mostly dark image with a bright horizontal bar across it at a certain depth as in Figure 3. Since the plane is at 45°, its vertical width is equal to the ST of the ultrasound beam at that depth. Measurements of the ST at multiple depths can be achieved by moving the transducer along the surface of the phantom.

3 Materials and Methods

Manufacturer	Scanner	Transducer	Linear (L), Curvilinear (C), or multi-row (M)	Rows	Elements per row	Nominal central frequency (MHz)
Siemens	S2000	4C1	C	1	128	2.75
Siemens	S2000	6C1	C	1	192	3.75
GE	Logiq E9	C1-5-D	C	1	192	3.5
GE	Logiq E9	9L-D	L	1	192	5
Siemens	S2000	18L6 HD	L	1	576	11.75
Siemens	S2000	14L5	M	3	192	9.5
Siemens	S2000	9L4	M	3	192	6.5
GE	Logiq E9	ML6-15-D	M	3	334	9.5

Table 1. Details of the 8 models of transducer studied. The nominal central frequency was defined as the midpoint of the frequency bandwidth reported in the manufacturers' manuals.

All scanners and transducers were retired from clinical use at the Royal Infirmary Edinburgh. A total of 8 transducer models were assessed, with 5 of them manufactured by Siemens (Erlangen, Germany) and the remaining 3 by GE (Boston, MA, USA). Details of each transducer are shown in Table 1. The procedure detailed in this section is often referred to as the ST method to differentiate from EPP measurements.

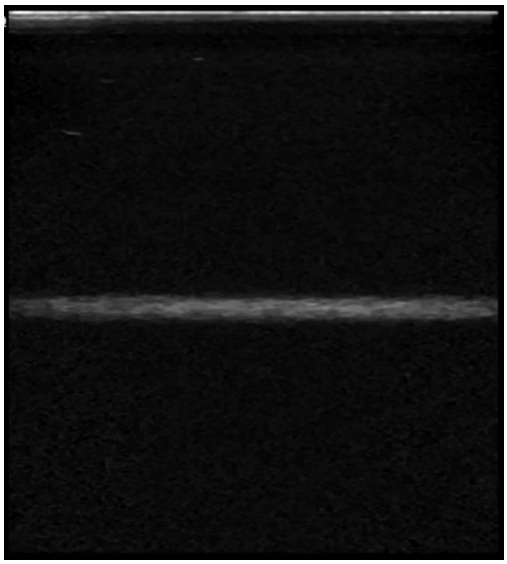


Figure 3. Ultrasound image of an ST phantom. The bright horizontal bar is the signal received from the 45° scattering plane, its vertical width is equivalent to the ST of the beam.

3.1 Lab procedure

A computer was connected to the scanner and set up to record the scanner's screen. An ATS 538NH beam profile and ST phantom (CIRS, Norfolk, Virginia, USA) was placed on a table next to the scanner (see Figure 4). One of the phantom's surfaces allows ST measurements over a depth range of 0 - 18 cm and the other over the range 8 - 22 cm. The former was preferred as it more closely matched the imaging ranges of the transducers. A rig that allowed smooth movement in the horizontal plane was placed on the table around the phantom, and the transducer was attached to it using claw clamps. The model of the transducer was looked up in the NHS records of EPP assessments to find the scanner settings used, and these were applied to the current scanner. Ultrasonic gel was applied to the phantom surface and the transducer was moved over the surface using the rig to ensure contact throughout its path. The transducer was moved to its starting position just beyond the point at which the deepest signal from the phantom's scattering plane could be discerned. A recording was begun on the computer and the rig's wheel was turned to move the transducer slowly across the phantom's surface toward the shallower end. The recording was

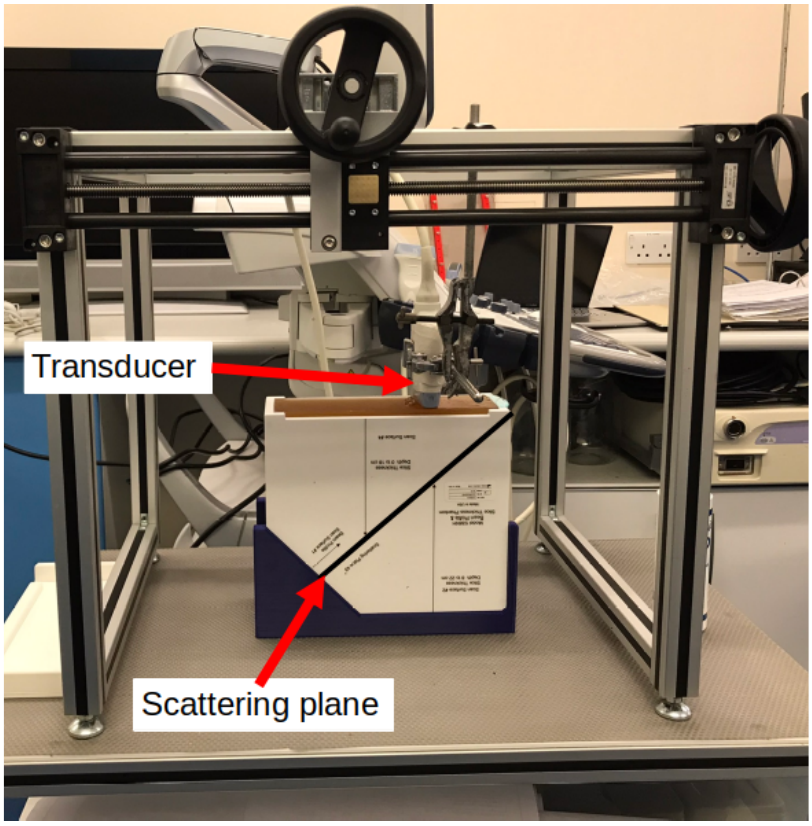


Figure 4. Setup used to obtain videos of the transducers' ST. The ST phantom is in the middle with its 45° scattering plane marked. The rig used to move the transducer smoothly across the phantom's surface is also visible.

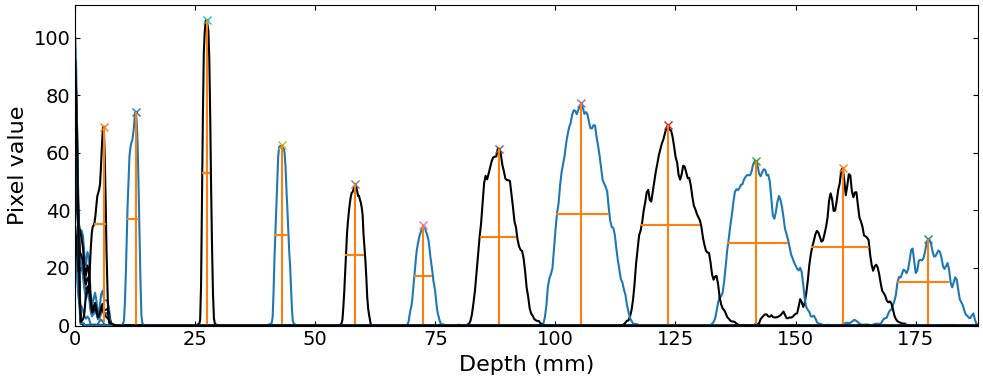


Figure 5. Series of superimposed pixel value profiles from two videos. The full width at half-maximum (FWHM) of each peak was taken to be the ST at that depth. Overlaid in orange are the peak FWHMs and heights as detected by the peak finding algorithm. Alternating black and blue distinguishes profiles from different frames in the video. Composed of data from two videos - one with shallow settings and one with deep.

ended once the desired depth had been traversed, and the video was saved. Each video then resembles Figure 3, with the bar slowly moving upwards while its vertical width changes.

Four videos were taken with each transducer: two with the scanner's settings adjusted for shallow visualisation, and two with settings for deep visualisation. In some cases, more than four videos were taken as settings (ones that had not been specified in the NHS records, such as TGC - time gain compensation) were adjusted further to improve the imaging. In total, 89 videos were taken with eight different transducer models.

3.2 Data analysis

3.2.1 Extracting slice thickness versus depth

To find how ST varied with depth in each video, they were analysed with a set of custom Python scripts [19], making extensive use of the OpenCV library [20]. The videos captured the entire scanner display, so each frame was cropped to include only the imaging region as in Figure 3. This was designated the region of interest (ROI) and its dimensions were recorded. Only every fifth frame was analysed as this did not strongly affect the quality of the results and it significantly reduced computation time. Each cropped frame was converted to greyscale using OpenCV's 'COLOR_BGR2GRAY' colour converter which preserves the colour channels' relative brightnesses for the human eye [21]. The frame was then averaged along the depth axis to form a profile of pixel value with depth. Background subtraction was performed using the shallow half of a profile taken when imaging the scattering plane at a deep point, spliced together with the deeper half of a profile taken when imaging a shallow point. This left the frame profiles as a mostly low signal with a single peak. The scipy.signal routine "find_peaks" was used to obtain the location, height and full width at half-maximum (FWHM) of each frame profile's peak measured in pixels. Figure 5 shows a subset of peaks

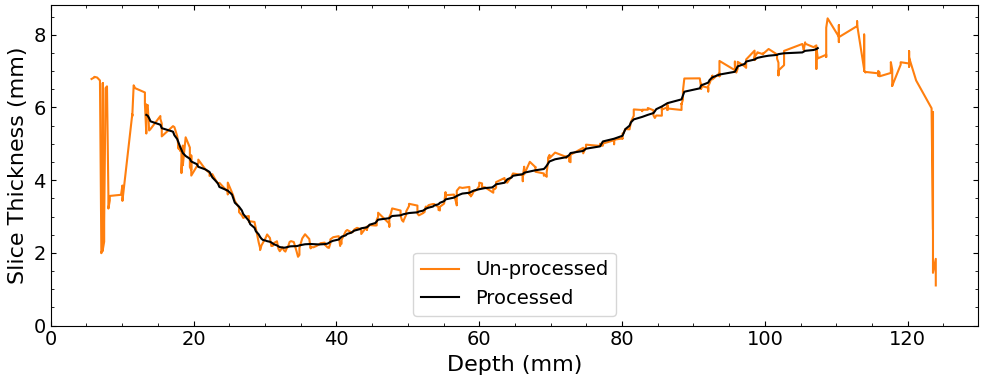


Figure 6. Representative example of pre- and post-processing ST depth profiles extracted from a video. Shallow and deep echoes have been trimmed in processing, and smoothing applied.

from two videos plotted together. The horizontal orange line represents the peak's FWHM which was taken to be the ST. These were converted to millimetres with a factor determined by equating the height of the ROI with the depth setting of the transducer in the video. The peak's width and depth were then multiplied by 0.94 to account for the lower speed of sound in the ST phantom ($1450\,\mathrm{m\,s^{-1}}$). The result was a set of plots showing how the ST varied with depth for each video; the orange line in Figure 6 serves as an example.

3.2.2 Profile processing

These plots were noisy and contained echoes at low and high depths, so further processing was applied. A pixel value of 20 was determined to be a reasonable threshold below which the signal was too faint and erratic to be reliably calculated by the peak finder, and therefore the profile was cut off at the last depth above this threshold. The low-contrast penetration (LCP) was taken to be the depth of the cut-off. In EPP measurements the LCP is defined as the depth at which speckle becomes lost in noise, or the point of deepest useful signal. Analysis of neighbouring frames sometimes yielded different STs (especially for faint signals), so the median was taken of all values recorded for the same depth. This was more appropriate than the average as it was less vulnerable to outlier measurements skewing the result. The profile was smoothed using scipy.ndimage's 'gaussian_filter1d' with the sigma parameter set to 3. Any decrease in ST at very shallow or very deep depths was trimmed as there is no physical mechanism for this so the cause must be an imaging artefact such as an echo. The black line in Figure 6 is an example of a processed curve.

3.2.3 Calculating the slice thickness resolution integral

At this point, the four ST profiles for each transducer (two obtained with shallow visualisation settings and two with deep) were combined to calculate the ST resolution integral - referred to as R (ST) to differentiate from R (EPP) measured with the EPP. A beam with a ST wider

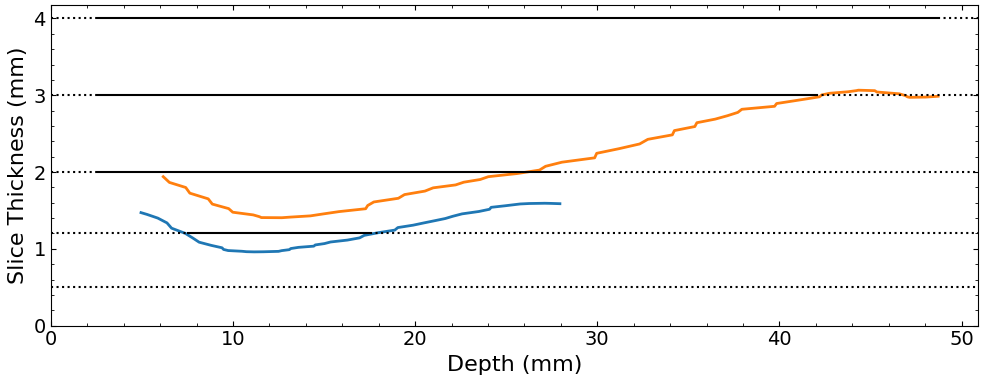


Figure 7. Example ST depth profiles with shallow (blue) and deep (orange) settings extracted from videos and processed. The horizontal lines show how the visualisation length L was obtained for each ST.

than an object was assumed to be unable to fully resolve that object due to out-of-plane signal polluting the image [22]. Therefore, the ST was taken to be equal to the minimum pipe diameter that could theoretically be imaged, in analogy to the EPP (see Section 2.3). The visualisation length, L, for a pipe could then be found by finding the depth range over which the ST was larger than that pipe's diameter. This process is shown in Figure 7, where the dotted horizontal lines show theoretical pipe diameters, and the solid sections denote the depth range over which the pipe would be visible given the above assumption. The length of each solid section is the value of L for that pipe diameter. The solid sections do not start from zero depth because of the dead zone of the transducer, the superficial region within which no physical signal is detected. Any diameter which was larger than the maximum ST had a visualisation length of the LCP minus the dead zone.

In contrast to the EPP where around 8 pipes of different diameters are imaged meaning interpolation between the points is necessary, there is no theoretical limit to the number of pipe diameters that can be sampled here. The calculated resolution integral became stable below 400 diameters, so this number was chosen. Any higher number would take longer to compute without offering any benefit.

A spreadsheet is currently used to calculate R, D_R and L_R for transducers assessed with the EPP. Since the number of diameters (pipes) measured with the videos was 400, manual data entry into this spreadsheet was not feasible. Therefore, a Python script was written to calculate R, D_R and L_R given pipe diameters and lengths, as described by Moran et al. [14]. The inverse of each pipe diameter was plotted against its L value to make an L-α plot as seen in Figure 2. The area under the curve was calculated using the trapezium rule (numpy.trapz), which is equal to the resolution integral, R. A diagonal line extending from zero towards the upper right corner of the plot was created, and scipy.optimize.minimize was used to minimise the difference between the areas above and below that line, thus finding the parameters L_R and D_R.

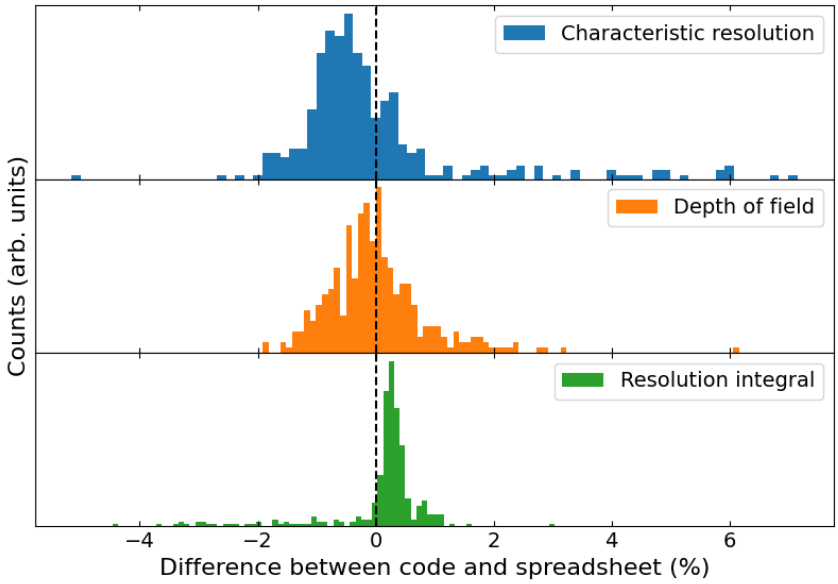


Figure 8. Histograms of the percentage difference between the Python code and the NHS spreadsheet for R, D_R and L_R. A positive difference means the code overestimated the value.

4 Results

4.1 Developmental results

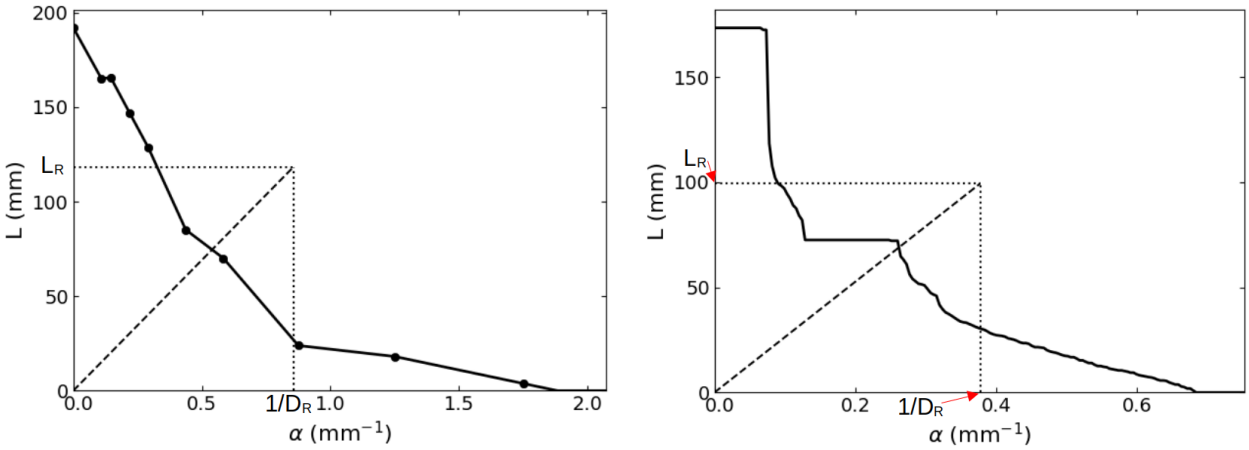
Intermediate steps in the analysis process produced some data worth exploring, these are presented in this section.

4.1.1 Slice thickness versus depth

The first useful result obtained from each video was the distribution of ST with depth in the phantom. Figures 6, 7 & 10 all contain examples of this data. The initial decrease to a minimum followed by steady growth to a wide peak is characteristic of all the ST plots obtained.

4.1.2 Performance of Python scripts

The Python code written to calculate R, D_R and L_R was validated against the currently-used spreadsheet with EPP measurements previously made by trained operators. 354 sets of EPP measurements and their spreadsheet-calculated results were tested. The resolution integrals calculated by the scripts were within 2.1% of the spreadsheet values with 95% confidence. The code was within 4% and 1.7% of the spreadsheet value for D_R and L_R respectively, also with 95% confidence. Histograms of the percentage differences are shown in Figure 8.



(a) (b)

Figure 9. Visualisation length L versus the inverse of (a) pipe diameter measured using the EPP, or (b) ST measured using the ST method. The data was taken from the same transducer, and the depths of field and characteristic resolutions are marked on the axes. The dotted rectangle is the L-α plot of an ideal beam with the same R. Individual data points are not shown for (b) as the line is approximately continuous.

There was a systematic bias towards overestimating R and underestimating D_R on the order of 0.4%, and instances of deviation up to nearly 8% could occur for D_R. All 354 sets of measurements were analysed in around one second.

4.1.3 L-alpha plots

One of the outputs of the code is a plot of visualisation lengths (L) versus α, the inverse of either the pipe diameter or ST. Representative examples of both are shown in Figures 9a & 9b. The shapes of the plots and the y-intercepts are broadly similar but the EPP plot intercepts the x-axis at a value about 3 times that of the ST plot. The ST curve varies more continuously but contains two major plateaus not present in the EPP data.

4.2 Changing focal depth

Settings on the ultrasound scanners can be adjusted to change the depth at which the transducer attempts to focus the beam, or even add multiple foci. For the 9L4 transducer, the ST depth profile was measured with the focus set at different depths while holding all other settings constant. The purpose of this was to find out how changing the focal depth affected the ST, and therefore how influential this could be on further findings. Figure 10 shows the results. For a focus depth setting of 10 mm, the minimum beam width was at a depth of 15 mm which then jumped to 31 mm when the focal depth setting was increased to 30 mm. Moving the focal depth beyond this tended to increase the ST at deep points, but the beam's thinnest point did not change significantly.

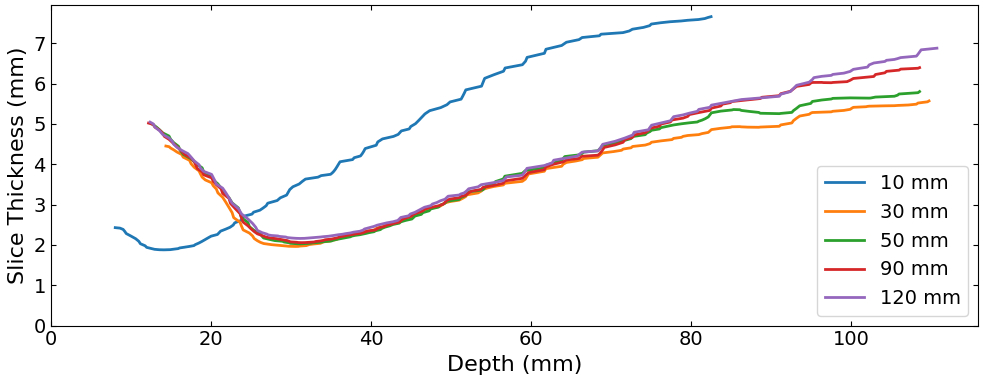


Figure 10. Slice thickness profiles for various focal depths. The legend shows the focal depth setting on the scanner, not the empirical focal depth. Data from the 9L4 transducer at frequency 4 MHz.

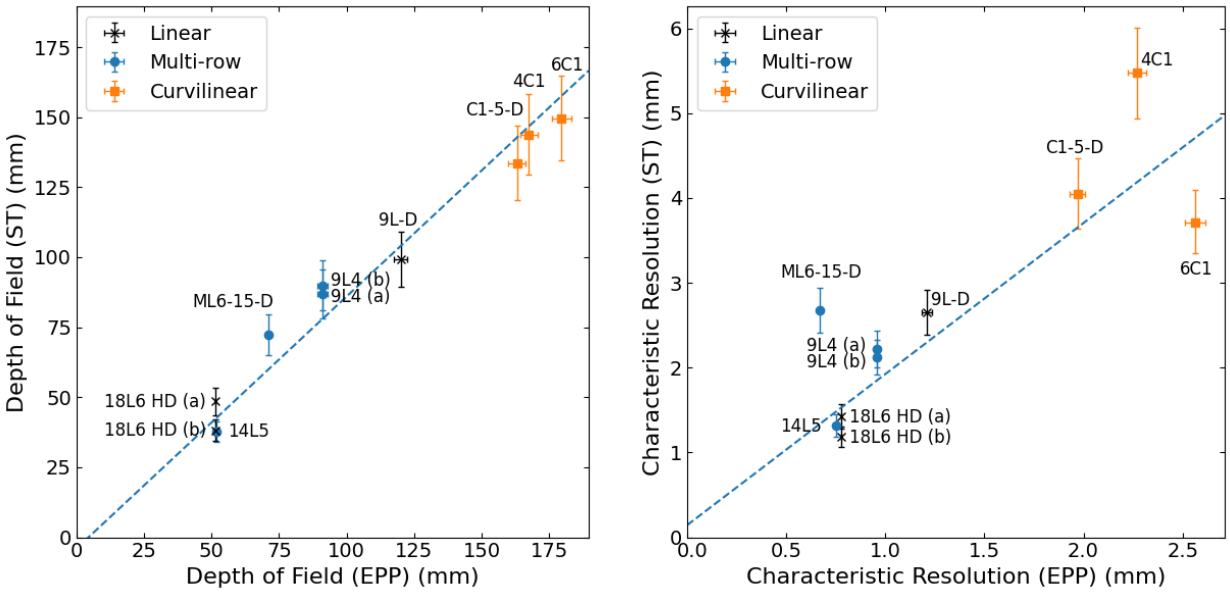
Transducer	R (EPP)	L_R (EPP) (mm)	D_R (EPP) (mm)	R (ST)	L_R (ST) (mm)	D_R (ST) (mm)
4C1	74	168	2.27	26	144	5.5
6C1	70	179	2.56	40	150	3.7
C1-5-D	88	163	1.97	33	134	4.1
9L-D	99	120	1.21	37	99	2.7
18L6 HD	66	51	0.78	33	43	1.31
14L5	70	52	0.75	29	38	1.31
9L4	94	91	0.96	41	88	2.2
ML6-15-D	105	71	0.67	27	72	2.7

Table 2. The resolution integrals R, depths of field L_R, and characteristic resolutions D_R of all transducers studied. Values are given as measured by the EPP and the ST method. The 18L6 HD and 9L4 results are averages from two individual transducers.

4.3 The slice thickness resolution integral

The resolution integral, characteristic resolutions, and depths of field measured using the ST method are listed in Table 2 along with the values obtained with the EPP. Uncertainties for EPP results were ±3% [4]. Uncertainty in ST results was ±10% on average and calculated from uncertainties in the method.

Figures 11a, 11b & 12 show these same results in graphical format, with R (ST), D_R (ST) and L_R (ST) plotted against their EPP counterparts. Correlation coefficients were calculated to be 0.18, 0.86, and 0.99 respectively. Linear fits were applied to the L_R and D_R plots, the fit parameters are given in the caption of Figure 11. Furthermore, both L_R (ST) and D_R (ST) negatively correlate with nominal central frequency: -0.95 and -0.86 respectively. See Appendix A for more details. There is no clear link between R calculated from ST and that from the EPP; the correlation coefficient is only 0.18. There are also no clear trends in terms of transducer type, with curvilinear, single-row, and multi-row transducers intermixed. The spread of R (ST) is much smaller than that of R (EPP).



(a) (b)

Figure 11. (a) Depth of field L_R calculated using the ST method versus L_R calculated with the EPP. A linear fit has been applied with gradient 0.90 ± 0.08 and y-intercept -4 ± 5.6 mm. (b) Characteristic resolution D_R calculated using the ST method versus D_R calculated with the EPP. A linear fit has been applied with gradient 1.8 ± 0.4 and y-intercept 0.14 ± 0.4 mm. Points are colour-coded by type of transducer. Any error bars smaller than the data points have been removed for clarity.

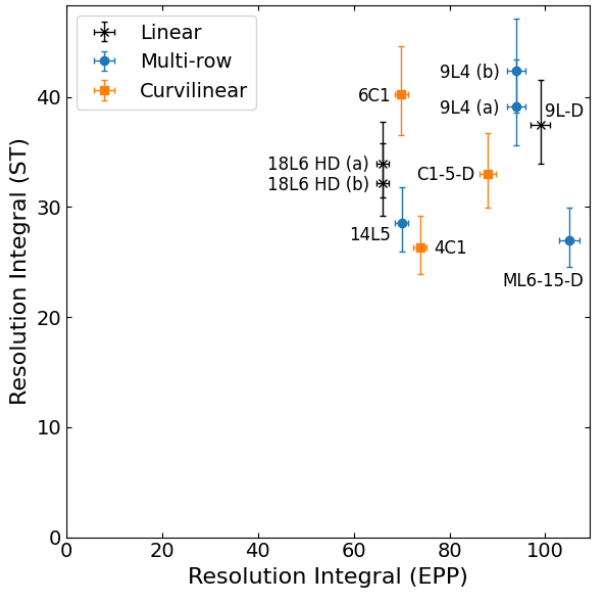


Figure 12. Resolution integral R calculated using the ST method versus R calculated with the EPP.

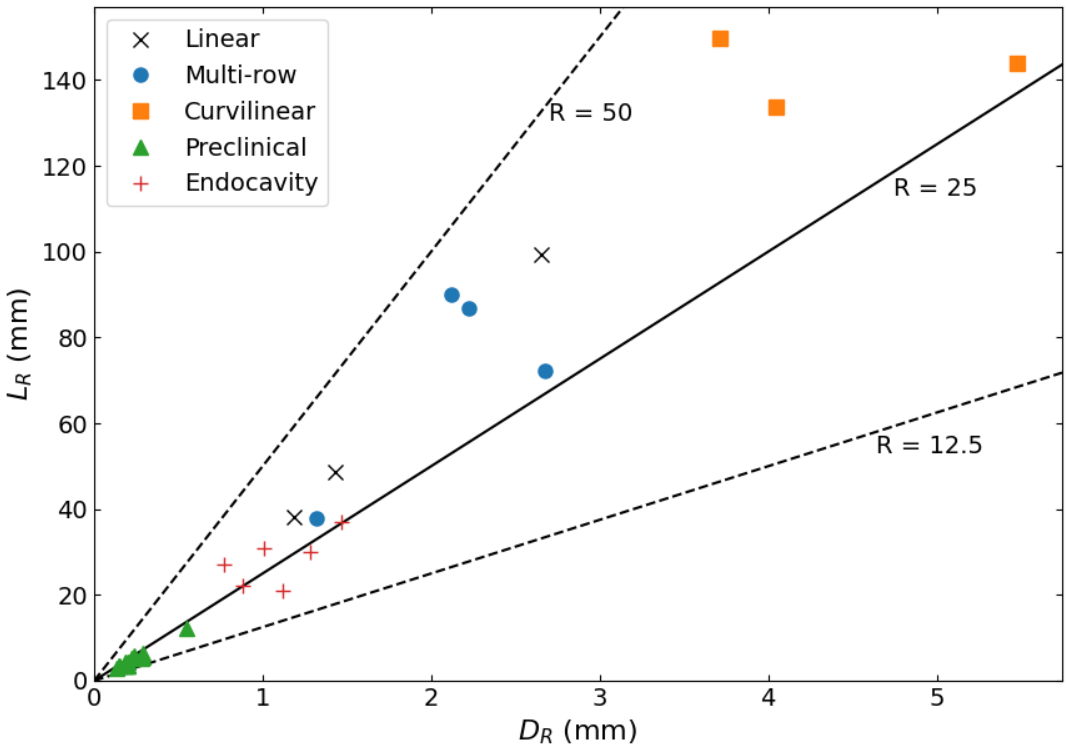


Figure 13. Depth of field plotted against characteristic resolution calculated using the ST method, classified into linear, multi-row, and curvilinear transducers. Also included are EPP results of preclinical and endocavity single-element transducers from [9] and [5]. The gradient of a line between a point and the origin is equal to the resolution integral, lines showing R = 12.5, 25, and 50 are plotted here.

4.4 Comparison to single-element transducers

Five of the transducers studied were only one element thick in the elevation direction, approximating, in this dimension, a single-element transducer. It was therefore tested whether their ST imaging performance was comparable to that of previously-reported results from single-element preclinical and endoscopic transducers [5, 9]. L_R is plotted against D_R in Figure 13 for all these transducers, the values from the other studies were obtained using the EPP. The preclinical transducers are clustered close to the origin, they all have R values above 12.5 and below 25. The endocavity transducers are located at higher L_R and D_R, and are scattered both above and below the R = 25 boundary. Two of the linear and one of the multi-row transducers are very close to this group while all three curvilinear transducers are located far in the top right. The rest of the linear and multi-row transducers are in between these groups, some closer to the R = 50 line.

5 Discussion

The outcomes of this research can offer insights into the effects of ST on the imaging performance of ultrasound systems. However, it is necessary to confirm the validity of the ST method before discussing the implications of results obtained using it.

5.1 Validation of the methodology

While the concept and implementation of the resolution integral have been rigorously tested, other parts of the methodology of this study do not have such established foundations. This section evaluates the developmental results with respect to literature and expectation with a view to raising the credibility of the findings.

An important part of the method involved the extraction of an ST depth profile from each video; an example was shown in Section 4.1.1 which is representative of the linear, multi-row, and curvilinear transducers studied. The profile's shape is linked to the focusing effect of the acoustic lens for the linear and curvilinear transducers, and the limited electronic focusing of the multi-row transducers, which cause the beam to converge in the near field and diverge in the far field. Typical beam shape in the ST plane has also been independently measured using a step change in backscatter [7], and with a needle hydrophone [23]. In both cases, very similar ST depth profiles were obtained.

Though this is promising and provides a solid foundation for the method, there was uncertainty in both the depth and FWHM of the ST peaks extracted from each video. The peak shape was variable, especially with very deep or shallow signals (see Figure 5), which caused uncertainty in the ST and the depth. Furthermore, the FWHM of a signal from a wire was taken as a measure of resolution [24], but using it to determine ST is arbitrary to an extent since ultrasound beams have no well-defined edge. Adding to this, the method assumed that the beam can only visualise a pipe with a diameter less than the ST. This is not necessarily the case, although the same assumption was made in a previous study [7]. In future, an exploration of these decisions could be undertaken, with corrections introduced if needed.

Once the videos had been analysed, a Python script calculated the resolution integral and associated parameters. The script had good accuracy and was significantly faster to use than the spreadsheet so it could save time and resources if used regularly. Typical deviations from the spreadsheet were mostly smaller than other uncertainties but the more extreme deviations could affect studies like this one with a small sample size. The script is also not user-friendly as it is run from the command line and would require knowledge of Python to use reliably. The script could be embedded into a website or basic Graphical User Interface to enhance accessibility, and an independent code review could potentially improve its accuracy.

Apart from calculating R, L_R, and D_R, the script generated L-α plots which were used to visualise the performance of each transducer (see Section 4.1.3). The plots generated from EPP and ST data were similar in shape, and they also resemble L-α plots from other studies

which suggests that the ST method is reasonable and captures the relevant physics [10, 25]. The smoothness of the ST profile eliminates the need for long-distance interpolation between data points and allows finer details to be resolved. However, these plots also reflect one of the main limitations of the method. Taking ST profiles with only the deepest and shallowest settings omits valuable information about the ST profile with intermediate settings. In some cases, a profile with intermediate settings would allow a given ST to be resolved over a greater depth range, but a lower L is reported instead. This lack of data is the cause of the sharp discontinuities present in all ST L-α plots including Figure 9b and leads to an underestimation of R. Taking more videos with settings optimised for intermediate depths could shrink or eliminate the discontinuities, with the drawback of more work being required to assess a single transducer.

The ST method is based on established techniques such as the resolution integral, and developmental results have broadly indicated agreement with the literature which is promising. While the method involves assumptions and is accompanied by large uncertainties, these assumptions are reasonable and future work could reduce many of the sources of uncertainty.

5.2 Changing the focal depth

An investigation into how changing the focal depth of a multi-row transducer affected its ST profile was undertaken to understand more about beam focusing in the elevation plane for synthesis with later results; the findings were presented in Section 4.2. The unique shift of the ST profile between focal depths of 10 and 30 mm (see Figure 10) was likely caused by the activation of only one row of transducer elements at a focal depth of 10 mm, but all three rows being active beyond this. Having a wide aperture produces a wide beam which then requires more extreme focusing to resolve superficial objects [11], so narrowing the aperture by turning off element rows may improve shallow imaging performance, explaining this aspect of the result. However, there are other aspects that are unexpected: firstly that there are only minor differences in the elevational beam shape between focus depths over 10 mm, and secondly that increasing the focus depth appears to increase the ST beyond a depth of 70 mm, thereby likely worsening the imaging performance. These points both suggest that the electronic focusing in the elevation plane may not be very effective. However, the impact this has on imaging performance is unclear as most focusing takes place in the lateral plane which is not visible here. An evaluation that includes all three dimensions is necessary to reveal more about the effect of ST on imaging performance.

5.3 The slice thickness resolution integral

Measurements with the EPP combine information from the lateral, axial, and elevational dimensions to assess the overall imaging performance of ultrasound systems. Comparing results from the EPP to those from the ST method enables a deeper understanding of the relationship between ST and imaging performance. The main findings of these comparisons were that there was a strong linear relationship between L_R (ST) and L_R (EPP) and that

both of these variables correlated with nominal central frequency. There was a less strong linear relationship between D_R (ST) and D_R (EPP), and a weaker correlation with central frequency. There was no simple relationship between R (ST) and R (EPP) but it was found that R (ST) ranged between 25% and 60% of R (EPP), similar to the resolution integrals of single-element transducers. It was often not possible to differentiate between multi-row and linear transducers using these measurements, but curvilinear transducers performed differently.

L_R in the ST plane was compared to L_R from the EPP because these quantities are analogous to the transducer focal region which plays a central role in imaging performance. The observed proportional relationship between L_R (ST) and L_R (EPP) was not unpredictable since the EPP combines data from all three dimensions including ST. An explanation for the correlation is that both L_R (EPP) and L_R (ST) also negatively correlate with nominal central frequency since lower frequencies experience less attenuation in tissue which increases range [22]. However, the ML6-15-D and 14L5 transducers have the same nominal central frequency but very different L_R values demonstrating that this alone cannot account for the correlation. A more complete explanation arises from aspects of transducer design. It is known that imaging performance is better within the focal region in both the lateral and elevation planes, making a substantial overlap of these regions advantageous [26]. If this overlap is maximised, L_R in both planes becomes similar leading to a strong correlation between the extent of a transducer's overall focal region and its elevational focal region. A linear fit was applied to the data and found that L_R (ST) is typically about 10% less than L_R (EPP). Transducers use electronic focusing in the lateral plane to keep the beam narrow over a long range whereas the elevation plane only has a cylindrical lens providing weak focusing or multiple rows of elements providing less extensive electronic focusing, which could account for this difference. Given these results, it is likely that ST has a significant effect on imaging performance since the elevational focal region is intimately and consistently linked with a transducer's overall focal region.

Like L_R, the characteristic resolution (D_R) was correlated between the ST plane and overall, but with a poorer linear fit quality and a lower correlation coefficient (0.86 versus 0.99). The linear fit parameters showed that D_R (ST) is typically around 1.8 times larger than D_R (EPP). This is likely because these transducers have superior beam focusing in the lateral plane, leading to better imaging when all three dimensions are included [27]. Again, a broadly linear relationship was expected since an improvement in resolution in the ST plane would also improve the transducer's overall resolution which is what the EPP measures. The idea that transducer frequency is responsible for the correlation is again contradicted by the ML6-15-D and 14L5 results which diverge for D_R (ST) despite identical nominal central frequencies. The design of the transducers offers a potential explanation for the poor quality of the linear fit. In terms of resolution, maximising imaging performance could be achieved by improving the lateral and elevational resolutions individually, rather than by making them similar as for L_R. This decoupling makes the lower correlation less surprising, as an improvement in lateral resolution would improve the overall resolution and hence D_R (EPP), but may not change the elevational resolution at all. As an example, the ML6-15-D transducer has a higher than expected D_R (ST) given its EPP value, but this could be because it benefits more in lateral resolution from having more elements per row than it

does in ST resolution from having three rows of elements. However, this predicts that the 18L6 HD transducer should have a greater D_R (ST) than the ML6-15-D (since it has more elements per row and only one row) which was not observed. This indicates that further confounding factors affect these results such as deterioration over time, variation between individual transducers, or differences in scanner settings. Furthermore, two of the curvilinear probes (4C1 and 6C1) also do not match expectations in ways that are not explained by the data in Table 2. All this suggests that, while elevational resolution does impact imaging performance, there are likely confounding factors present that are not accounted for in this analysis.

The resolution integral, R, represents the overall transducer imaging performance. The key result was that no clear relationship between R (ST) and R (EPP) was found, despite the fact that an improved imaging performance in the elevation plane should logically result in an improvement overall. It is possible that the same confounding factors affecting the measurements of D_R could have erased any underlying relationship. This is likely because the same transducers which were outliers in D_R (Figure 11b) are again those that are furthest from showing a positive correlation. Taking the ratio of L_R and D_R could amplify the differences, and the large uncertainties involved ($\pm 10\%$) could further complicate any potential relationship. However, it is not unexpected to find that a transducer's overall imaging performance is not a simple function of its performance in the elevation plane, especially since most of the focusing technology is tailored to the lateral plane. This prioritisation of the lateral plane also explains why R (ST) values were only 25-60% of R (EPP). All this suggests that ST imaging performance is an important limiting factor on overall imaging performance, a result which reflects previous findings that R values independent of ST were over 2.5 times higher than when ST was included [16].

Since the imaging performance of a transducer is dependent on the application it was designed for, being able to distinguish between transducer applications with the ST method is an argument in favour of ST having a strong impact on imaging performance. However, no trends were found between R (ST) and R (EPP), even when taking transducer type into account. While the resolution integral does allow some differentiation between transducer applications, there is significant overlap in these categories, and L_R and D_R must be used to fully differentiate [10]. This could explain why no transducers could be reliably classified using R. When L_R and D_R were plotted it became possible to distinguish curvilinear probes, but not linear or multi-row. The curvilinear transducers had the lowest nominal central frequencies, which helps separate out transducers by application [10] so this may not have been due to the influence of ST. However, the reason for the similarity could be because, with the shallowest settings, only the central row of elements in a multi-row was activated, effectively simulating a linear transducer. Adding to this, the focusing ability did not seem to be very strong in the elevation plane as the ST profile did not change much as focus depth was increased. Videos were only taken with shallow and deep settings, therefore all of the data collected was vulnerable to these effects, so it is easier to see how the difference between linear and multi-row transducers could be underestimated. Therefore, it is still likely that ST plays a significant part in ultrasound imaging performance even though the ST method is not sensitive enough to sort transducers by application.

The outcomes of this study were affected by several limitations. Firstly, only 10 individual transducers and 8 models of transducer were assessed, limiting the generalisability of the results. Also, differences between the EPP and ST phantom could have affected the results. The EPP is made of agar-based TMM [28] while the ST phantom is made of urethane [29]. A correction was made for the speed of sound, but differences in image construction can arise [30, 31]. Furthermore, the EPP measurements were made by finding the point at which anechoic pipes were no longer visible - a degradation in contrast, while ST measurements were based on a peak's width - a high contrast region. This means the ST measurements were based on sharper boundaries which may offer less uncertainty, but the EPP measurements are more clinically relevant since they were based on the human eye. In addition, the LCP was defined differently for the EPP and ST method. LCP is "a measure of the depth at which the speckle signal from the background scattering material disappears or is lost in electronic noise" [32]. However, for the ST method, the LCP was determined by the depth at which the peak finder could no longer calculate the ST or the depth at which the highest pixel value in the peak fell below 20, whichever was shallower. These should be reasonably similar, as the cut-off pixel value of 20 was chosen because a signal less than this could not be discerned by eye, however, the methods are different which may have affected the ST results. The uncertainty in each R (ST) value is on the order of $\pm 10\%$ which is large. Also, quantities measured with the EPP were generally obtained more than 8 years ago, and not for the same individual transducers as studied here. It is known that transducers can deteriorate over time and that this changes their imaging performance as measured with the EPP [5]. Furthermore, values obtained with different physical transducers of the same model were different (see Figures 12, 11a and 11b). Therefore, the reported EPP values may not be accurate for the transducers as they were assessed here, and this could have affected the results.

6 Conclusions

The procedure detailed in this report was based on established techniques and reasonable assumptions while obtaining interesting results which are relevant to assessing ultrasound systems. It was found that slice thickness plays an important role in ultrasound imaging performance, but the nature of the relationship is not clear. A larger elevational focal region is strongly associated with a larger overall focal region. However, the resolution within that region is affected by confounding factors most likely arising from the lateral plane. It is possible to build up a reasonably complete picture of a transducer's imaging performance just by considering its ST, but high uncertainties and a small sample size limited the strength of conclusions that could be drawn from these results. More research is needed to understand exactly how slice thickness influences imaging performance.

There are a number of avenues of investigation to explore in future. Only 10 transducers and 8 different models were assessed for this study, including more transducers and models would strengthen any conclusions. Taking videos with settings optimised for intermediate imaging would make the results more accurate. Making EPP measurements concurrently

with ST videos with the same transducer would eliminate any concern about deterioration over time and differences between transducers. It would have been very useful to have more data on beam focusing in the lateral plane as this could have been used to explain the differences between D_R (EPP) and D_R (ST), so future studies could involve an individual assessment of lateral focusing and resolution as well. The measurement of how the ST profile changed with focal depth was only undertaken for one transducer meaning that the results are not generalisable. Assessing more transducers, including linear and curvilinear, would enable a better understanding of beam focusing in the elevation plane. An in-depth review of the Python code for calculating the resolution integral could reduce the difference with the spreadsheet, and packaging it into a user-friendly application could allow faster evaluation of transducers.

References

[1]R. Smith-Bindman et al., "Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000-2016", JAMA **322**, 843–856 (2019).

[2]J. M. Thijssen, G. Weijers, and C. L. de Korte, "Objective performance testing and quality assurance of medical ultrasound equipment", Ultrasound in Medicine & Biology **33**, 460–471 (2007).

[3]A. Shaw and R. Hekkenberg, *Standards to support performance evaluation for diagnostic ultrasound imaging equipment*, NPL Report AC 2, 2007.

[4]S. D. Pye, W. Ellis, and T. MacGillivray, "Medical ultrasound: a new metric of performance for grey-scale imaging", Journal of Physics: Conference Series **1**, 187 (2004).

[5]C. M. Moran et al., "The imaging performance of preclinical ultrasound scanners using the Edinburgh Pipe Phantom", Frontiers in Physics **10**, 802588 (2022).

[6]T. MacGillivray, W. Ellis, and S. D. Pye, "The resolution integral: visual and computational approaches to characterising ultrasound images", Physics in Medicine & Biology **55**, 5067–5088 (2010).

[7]D. E. Rowland et al., "The automated assessment of ultrasound scanner lateral and slice thickness resolution: use of the step response", Ultrasound in Medicine & Biology **35**, 1525–1534 (2009).

[8]C. McLeod et al., "Evaluation of intravascular ultrasound catheter-based transducers using the resolution integral", Ultrasound in Medicine & Biology **44**, 2802–2812 (2018).

[9]S. Inglis et al., "Assessing the imaging capabilities of radial mechanical and electronic echo-endoscopes using the resolution integral", Ultrasound in Medicine & Biology **40**, 1896–1907 (2014).

[10]C. M. Moran et al., "The imaging performance of diagnostic ultrasound scanners using the Edinburgh Pipe Phantom to measure the resolution integral - 15 years of experience", Ultraschall in Med **43**, 393–402 (2020).

[11]J. T. Bushberg and J. M. Boone, *The essential physics of medical imaging* (Lippincott, Williams & Wilkins, 2011).

[12]P. N. T. Wells, "Absorption and dispersion of ultrasound in biological tissue", Ultrasound in Medicine & Biology **1**, 369–376 (1975).

[13]W. R. Hendee and E. R. Ritenour, *Ultrasound transducers* (John Wiley & Sons, Ltd, 2002) Chap. 20, pp. 317–330.

[14]C. M. Moran, S. Inglis, and S. D. Pye, "The resolution integral – a tool for characterising the performance of diagnostic ultrasound scanners", Ultrasound **22**, 37–43 (2014).

[15]E. L. Madsen, "Quality assurance for grey-scale imaging", Ultrasound in Medicine & Biology **26**, S48–S50 (2000).

[16]H. Carstairs et al., "A novel ultrasound phantom to quantify the effect of slice thickness on imaging performance", 2019 IEEE International Ultrasonics Symposium, Glasgow, Scotland, 2404–2407 (2019).

[17]A. Goldstein, "Slice thickness measurements", Journal of Ultrasound in Medicine **7**, 487–498 (1988).

[18]M. L. Skolnick, "Estimation of ultrasound beam width in the elevation (section thickness) plane", Radiology **180**, 286–288 (1991).

[19]D. Roddy, *MPhys project code*, `https://github.com/roddydr123/ultrasound`, [Online; accessed 20-03-23], 2023.

[20]G. Bradski, "The OpenCV Library", Dr. Dobb's Journal of Software Tools **25** (2000).

[21]C. Poynton, *Frequently asked questions about color*, `http://poynton.ca/PDFs/ColorFAQ.pdf`, [Online; accessed 20-03-23], 1997.

[22]C. R. Hill, J. C. Bamber, and G. R. ter Haar, *Physical principles of medical ultrasonics* (John Wiley & Sons, Ltd, 2004).

[23]J. A. Brown et al., "Fabrication and performance of a 40-MHz linear array based on a 1-3 composite with geometric elevation focusing", IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **54**, 1888–1894 (2007).

[24]N. M. Gibson, N. J. Dudley, and K. Griffith, "A computerised quality control testing system for B-mode ultrasound", Ultrasound in Medicine & Biology **27**, 1697–1711 (2001).

[25]S. D. Pye and W. Ellis, "The resolution integral as a metric of performance for diagnostic grey-scale imaging", Journal of Physics: Conference Series **279**, 012009 (2011).

[26]T. Szabo, *Diagnostic ultrasound imaging: inside out* (Elsevier Science & Technology, 2013).

[27]P. R. Hoskins, K. Martin, and A. Thrush, eds., *Diagnostic ultrasound : physics and equipment* (Cambridge University Press, Cambridge, 2010).

[28]K. V. Ramnarine, T. Anderson, and P. R. Hoskins, "Construction and geometric stability of physiological flow rate wall-less stenosis phantoms", Ultrasound in Medicine & Biology **27**, 245–250 (2001).

[29]CIRS, *Model ATS 538NH beam profile & slice thickness phantom user guide*, `https://www.cirsinc.com/wp-content/uploads/2020/12/ATS538NH-UG-073120.pdf`, [Online; accessed 21-03-23], 2020.

[30]A. Goldstein, "The effect of acoustic velocity on phantom measurements", Ultrasound in Medicine & Biology **26**, 1133–1143 (2000).

[31]H. L. Green, N. J. Dudley, and N. M. Gibson, "A comparison between urethane and gel test objects for ultrasound imaging performance testing", Ultrasound **15**, 105–108 (2007).

[32]N. J. Dudley and N. M. Gibson, "Is grey level a suitable alternative to low-contrast penetration as a serial measure of sensitivity in computerised ultrasound quality assurance?", Ultrasound in Medicine & Biology **43**, 541–545 (2017).

Appendices

Note — Appendices are provided for completeness only and any content included in them will be disregarded for the purposes of assessment.

A Correlation matrix

Figure 14. Correlation matrix of all calculated R, D_R and L_R, as well as transducer nominal central frequency.