OBJECTIVE VS SUBJECTIVE MEASURE OF THE LOW CONTRAST PENETRATION TEST IN ULTRASOUND IMAGING

D. R. Hewer [1], M. L. Camilleri [1], M. Busuttil [1], K. Vella [1], A. Pule [1], K. A. Micallef [1]

[1] Department of Medical Physics, Faculty of Health Sciences, University of Malta, Malta

Abstract – **This study compares subjective and objective approaches for measuring low contrast penetration (LCP) in ultrasound (US) quality assurance (QA). Thirty repeated phantom tests were conducted using two linear, and one curvilinear probe. An objective single phantom image (SPI) method determined the depth of penetration from signal-to-noise ratio (SNR) curves using a polynomial fit and defined thresholds. SPI measurements showed markedly lower variability (standard deviation (SD) < 2 mm) compared to manual measurements (SD up to 9 mm). Agreement between methods was strongest for the curvilinear probe. The findings support integrating objective LCP analysis into routine US QA programmes.**

Keywords – **ultrasound, low contrast penetration, automated methods, quality assurance.**

I. INTRODUCTION

Low contrast penetration (LCP) testing evaluates an ultrasound (US) system's sensitivity by measuring the depth at which speckle signals are distinguishable from the background scattering material in a phantom [1]. It is a critical parameter for identifying gradual performance degradation over time, often preceding visible artefacts or system failure [2,3]. According to professional standards such as IPEM Report 102 [4] and AAPM Report No. 1 [5], LCP assessment is typically assessed visually by measuring the depth at which the last discernible structure is observed. However, this subjective approach introduces inter- and intra-observer variability, dependent on experience, visual acuity, and ambient conditions such as display brightness or room lighting.

To mitigate this subjectivity, objective computational methods have been proposed, with Gorny et al. (2005) [6] introducing the single phantom image (SPI) approach for deriving depth of penetration (DOP) from quantitative image analysis. The SPI method computes signal-to-noise ratio (SNR) profiles across depth and defines the penetration limit as the depth at which the SNR falls below a predefined threshold. Despite its promise, few studies have compared SPI-derived DOP values against manual measurements in routine quality assurance (QA) settings.

This study aims to evaluate the consistency and reliability of the SPI method against traditional visual estimation for different US probe types, thus determining its suitability as a standardised, observer-independent measure of LCP.

II. MATERIALS AND METHODS

Data Acquisition

US uniformity images were acquired from thirty repeated US phantom scans performed using the same single diagnostic system at the Medical Imaging Department of Mater Dei Hospital, Malta. Three probes were assessed: i18LX5, i24LX8 (linear), and i8CX1 (curvilinear). Each scan maintained consistent imaging parameters, including depth, focus, and gain settings, while minor variations in time-gain compensation (TGC) were recorded. For the curvilinear probe, the maximum analysis depth was limited to 170 mm to exclude reverberation artefacts from the phantom base.

Manual Method

Five medical physicists (MPs) (MP1–MP5) independently reviewed stored images and recorded the DOP at which the final visible structure could be identified. Observer experience varied from <1 year to >8 years, providing a broad representation of practical variability.

Objective Method

The SPI analysis was implemented in Python. For each test, the SNR was computed as the ratio of mean pixel value (MPV) within the phantom to the background noise from an in-air region, using a vertical line profile across image depth (Eq 1) [6]. The tail region of the SNR curve, beginning after the maximum signal, was fitted using a second-degree polynomial. The DOP was then determined as the depth where the fitted curve intersected the SNR threshold (4.5 for linear, 1.4 for curvilinear probes), as described by Gorny et al. (2005) [6]. Figure 1 shows an example of the curve-fitting process, for the i18LX5 probe. Equation 1 shows the calculation for the SPI method.

$$SNR(d) = \mu_1(d) \div \mu_a(d) \quad (1)$$

where $\mu_1(d)$ is the MPV for the phantom image, and $\mu_a(d)$ is the MPV of the in-air image, both at depth d.

Statistical Analysis

Descriptive statistics were computed for both manual and SPI-derived DOPs. Standard deviation (SD) was used as a measure of repeatability. Inter-observer variability was evaluated, and comparisons between objective and subjective methods were analysed across probe types.

Software-related uncertainty arising from ImageJ-based measurements was estimated at ±0.001 mm.

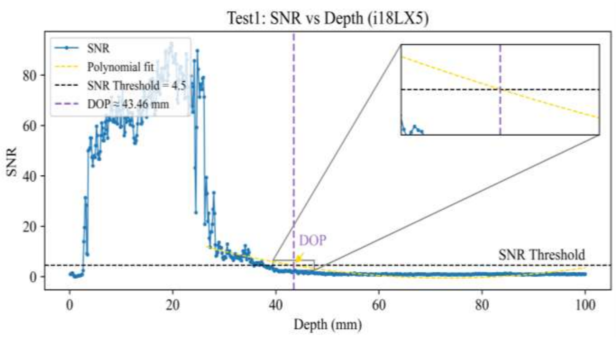

Fig. 1 Graph of SNR vs Depth for i18LX5 probe (test 1)

III. RESULTS

The SPI method demonstrated markedly lower variability across all probe types, with SD below 2 mm compared to manual values that reached SD of 6-9 mm, especially among less experienced observers. For the linear probes (i18LX5, i24LX8), SPI-derived DOPs were consistently lower (~30-50 mm) than manual readings (~60-70 mm), reflecting the stricter SNR threshold criterion. In contrast, for the curvilinear probe (i8CX1), agreement between manual and SPI DOPs was stronger (130-165 mm), aligning with expectations for lower-frequency transducers that exhibit slower signal attenuation. Less experienced MPs (MP5) yielded inconsistent readings, further illustrating the subjectivity of visual interpretation. Table 1 shows the automated and manual DOP measurements, performed by each MP, for the i8CX1 probe.

Table 1: i8CX1 curvilinear probe automated and manual DOP (mm) measurements

Test no.	SPI	MP1	MP2	MP3	MP4	MP5
1	138.06	157.35	150.11	147.91	157.62	144.00
2	141.83	156.40	162.88	145.13	152.04	150.35
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	144.05	162.58	154.27	141.65	155.29	142.44
Mean	145.89	158.20	157.46	143.36	154.36	146.66
SD	3.75	2.99	3.80	2.39	2.96	12.54

Figure 2 shows a plot of both manual and automated DOP measurements for each test, for the i8CX1 probe.

Similarly, tables 2 and 3 show the automated and manual DOP measurements, for the linear i18LX5 and i24LX8 probes respectively, while figures 3 and 4 display plots of both manual and automated DOP measurements for each test, for the i18LX5 and i24LX8 probes respectively.

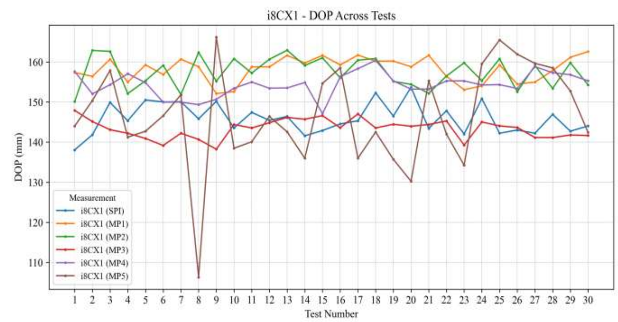

Fig. 2 DOP measurements for each test for the i8CX1 probe

Table 2: i18LX5 linear probe automated and manual DOP (mm) measurements

Test no.	SPI	MP1	MP2	MP3	MP4	MP5
1	43.46	53.48	49.13	64.02	52.41	52.26
2	46.45	52.80	59.56	67.79	58.85	64.22
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	46.56	51.44	51.51	64.08	52.32	63.74
Mean	46.45	56.99	54.55	66.93	54.73	64.09
SD	1.88	2.67	3.10	1.13	2.61	7.73

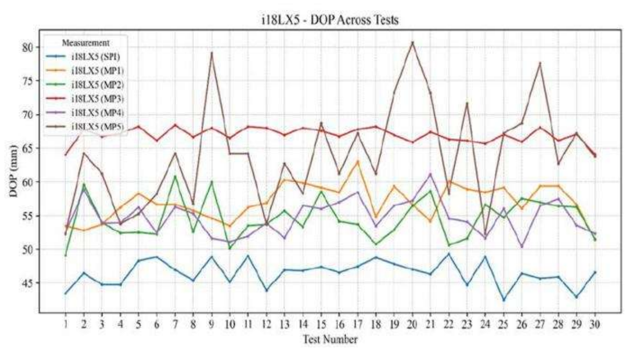

Fig. 3 DOP measurements for each test for the i18LX5 probe

Table 3: i24LX8 linear probe automated and manual DOP (mm) measurements

Test no.	SPI	MP1	MP2	MP3	MP4	MP5
1	29.50	56.79	58.27	54.87	58.18	78.03
2	30.18	59.49	46.47	54.77	58.41	68.59
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	29.86	60.91	55.13	55.49	57.60	61.12
Mean	30.22	57.09	54.44	55.10	55.66	42.22
SD	0.89	1.63	3.82	1.21	2.83	11.69

Table 4 demonstrates the coefficient of variance (COV) produced by both methods, for each probe. Equation 2 shows the calculation for the COV.

$$COV = (\sigma \div \mu) \times 100\% \quad (2)$$

where σ is the standard deviation, and μ is the mean, of recorded DOP measurements, for each probe.

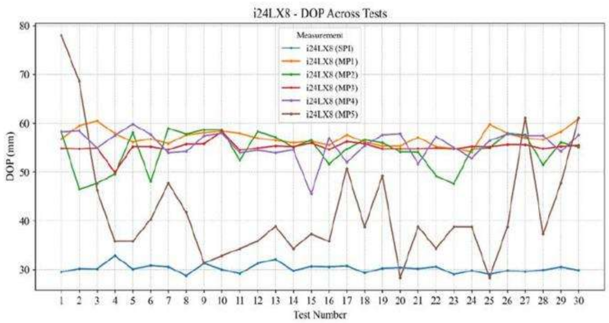

Fig. 4 DOP measurements for each test for the i24LX8 probe

Table 4: COV (%) for each probe

	i18LX5	i24LX8	i8CX1
SPI	4.04	2.96	2.57
MP1	4.68	2.85	1.89
MP2	5.68	7.10	2.41
MP3	1.69	2.20	1.67
MP4	4.77	5.08	1.92
MP5	12.06	27.68	8.55

IV. DISCUSSION

The findings confirm that the SPI method provides a more stable and repeatable measure of LCP than manual assessment. Manual readings demonstrated significant inter-observer variability, primarily linked to experience level and visual acuity of faint echoes. Less experienced participants tended to overestimate DOP by identifying faint residual patterns below the noise threshold.

In the SPI method, DOP underestimation in high-frequency linear probes arises from steeper attenuation and the use of a second-degree polynomial fit, which may not optimally capture the complex decay of the SNR tail. Alternative curve-fitting approaches, such as exponential or spline models, could better model the asymptotic decline of the signal. The TGC slider positions were noted, to be replicated for each image, to achieve consistency. However, given that the images were taken on different consecutive days, each time the TGC were readjusted to the initial positions recorded, which may have contributed to some variations. Therefore, reinforcing the need for a fixed TGC preset for routine QA to avoid any variations.

Although the SPI implementation employed a simplified vertical line profile, it effectively met the purpose of the LCP test, which is to monitor consistency across repeated acquisitions rather than determining absolute imaging limits. The low SD values indicate that the SPI method is well suited for longitudinal QA applications.

The results are consistent with prior findings by Gorny et al. (2005) [6], Gibson et al. (2001) [2], and Fiori et al. (2023) [7], which similarly emphasise the value of objective SNR-based assessment in maintaining QA reliability. While the SPI method's simplicity may limit precision at significant depths, its automation potential and reproducibility make it a promising tool for LCP evaluations.

V. CONCLUSIONS

The objective SPI method demonstrated superior repeatability and lower measurement variability than subjective visual assessment across multiple US probes. Agreement with manual results was strongest in the curvilinear probe, where attenuation effects are less pronounced. Differences between methods were accentuated in high-frequency probes, underscoring the need for consistent system settings. The study supports integration of SPI-based analysis into US QA protocols to ensure consistent, and observer-independent evaluation of imaging depth performance.

VI. RECOMMENDATIONS

To enhance precision, future research should:

- Explore higher-order or adaptive curve-fitting models.
- Incorporate machine learning approaches to automate threshold detection and region of interest (ROI) selection.
- Standardise imaging settings, including TGC, across QA sessions.
- Expand the study across multiple US systems and phantom types to validate generalisability.

VII. ACKNOWLEDGMENT

The author thanks the supervisory team, the Medical Imaging Department of Mater Dei Hospital, and all participating medical physicists for their contribution to data collection and analysis.

VIII. REFERENCES

1. Dudley, N. J. and Gibson, N. M. (2017). 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(2):541–545.
2. Gibson, N. M., Dudley, N. J., and Griffith, K. (2001). A computerised quality control testing system for b-mode ultrasound. Ultrasound in medicine & biology, 27(12):1697–1711.
3. Hangiandreou, N. J., Stekel, S. F., Tradup, D. J., Gorny, K. R., and King, D. M. (2011). Four-year experience with a clinical ultrasound quality control program. Ultrasound in medicine & biology, 37(8):1350–1357.

4. Grazhdani, H., David, E., Ventura Spagnolo, O., Buemi, F., Perri, A., Orsogna, N., Gigli, S., and Chimenz, R. (2018). Quality assurance of ultrasound systems: current status and review of literature. Journal of ultrasound, 21(3):173–182.
5. Goodsitt, M. M., Carson, P. L., Witt, S., Hykes, D. L., and Kofler, J. M. (1998). Real-time B-mode ultrasound quality control test procedures.
6. Gorny, K. R., Tradup, D. J., and Hangiandreou, N. J. (2005). Implementation and validation of three auto mated methods for measuring ultrasound maximum depth of penetration: application to ultrasound quality control. Medical physics, 32(8):2615–2628.
7. Fiori, G., Fuiano, F., Schmid, M., Conforto, S., Sciuto, S. A., and Scorza, A. (2023). A comparative study on depth of penetration measurements in diagnostic ultrasounds through the adaptive SNR threshold method. IEEE Transactions on Instrumentation and Measurement, 72:1–8.

Contacts of the corresponding author:

Author: Dylan Robert Hewer
Institute: University of Malta
Street: Tal-Qroqq
City: Msida
Country: Malta
Email: dylanhewer.dh@gmail.com