

Article

A Preliminary Study on the Adaptive SNR Threshold Method for Depth of Penetration Measurements in Diagnostic Ultrasounds

Giorgia Fiori ^{1,*}, Fabio Fuiano ¹, Andrea Scorza ¹, Jan Galo ², Silvia Conforto ¹ and Salvatore Andrea Sciuto ¹

¹ Engineering Department, Roma Tre University, via della Vasca Navale 79, 00146 Roma, Italy; fabio.fuiano@uniroma3.it (F.F.); andrea.scorza@uniroma3.it (A.S.); silvia.conforto@uniroma3.it (S.C.); salvatore.sciuto@uniroma3.it (S.A.S.)

² Clinical Engineering Service, IRCCS Children Hospital Bambino Gesù, Piazza di Sant'Onofrio 4, 00165 Roma, Italy; jan.galo@opbg.net

* Correspondence: giorgia.fiori@uniroma3.it

Received: 29 July 2020; Accepted: 16 September 2020; Published: 18 September 2020



Featured Application: The present work is aimed at providing a novel image analysis-based method for the maximum depth of penetration (DOP) measurement in Quality Assessment (QA) of medical ultrasound systems.

Abstract: Maximum depth of penetration (DOP) is among the most relevant parameters in quality assurance programs for Ultrasound (US) scanners. Nowadays, a generally-accepted protocol for DOP estimation is still awaited and, in common practice, DOP is visually assessed despite the low accuracy. To overcome the eye-based assessment subjectivity, automatic image analysis methods have been proposed in literature. The present work focuses on a novel automatic method, namely the adaptive Signal to Noise Ratio (SNR) threshold method (AdSTM), developed in the MATLAB environment, by comparing it with an existing automatic approach, namely the tangent threshold method (TTM), and the mean judgment of eight observers (naked eye method). The three investigated methods were applied on data acquired from four US scanners for general purpose imaging, equipped with linear, convex, and vector array probes. Tests were carried out in two different configuration settings (raw scanner and default preset working conditions). AdSTM outcomes were tested by means of Monte Carlo Simulations. Most of measurement results were compatible despite the fact that the AdSTM seemed to be more sensitive and faster than the TTM. The results analysis confirms the higher dispersion of the naked eye method in DOP assessment with respect to the proposed automatic methods.

Keywords: depth of penetration; adaptive SNR threshold method; tangent threshold method; quality assurance programs; US scanner

1. Introduction

As pointed out in several studies [1–9], in recent years quality assurance (QA) programs, able to objectively provide ultrasound (US) parameter measurements, have become more and more important in the assessment of US scanner functionality, given that they are implied in a wide range of well-known clinical applications for the visualization of anatomical districts worldwide. It is therefore of primary importance to make sure that clinicians can identify and characterize possible abnormalities in an anatomic image (e.g., malignancies). This should be possible at any depth with the highest reliability, resulting in an improvement of both diagnostic and therapeutic phases and in a reduction of the

clinician mismatching to the minimum as well as patient discomfort due to wrong therapeutic choices. Such results can be achieved as long as the US scanners are able to provide a consistent quality in the images after having checked several B-mode US parameters. Among the main ones, maximum depth of penetration (DOP) deserves particular attention—defined as the maximum low contrast penetration depth at which US speckle due to the scattering within a tissue-mimicking material (TMM), is displayed by the US scanner without noise [1,2,4,8,10,11]. In particular, speckle pattern is prominently displayed in the image when considering depths much smaller than DOP, while the situation changes at greater depths than DOP where the electronic noise is predominant with respect to speckle. The DOP is an indicator of how well the US probe is able to display the features of anatomical districts far from the probe–skin/phantom interface. Browne et al. suggested that it could be considered a predictor for the clinician’s image quality perception [12]. Nowadays, there is neither a wide consensus nor a commonly-accepted method for DOP estimation and in many cases it is still visually assessed [13,14]. However, it should be highlighted that DOP operator dependence is detrimental for assessment objectivity. On the other hand Mannila et al. [15], after having defined DOP as the depth where speckle starts to be dominated from noise, pointed out how its estimation is necessary because of its influence on vertical distance accuracy measurements. Another scientific study [16] provided an alternative definition of such parameters as the maximum depth at which speckle is still visible. The estimation was based on the standard deviation computation of both speckle and noise, from two consecutive images captured by holding the probe still. Nevertheless, image analysis-based off-line methods have been proposed in literature [10,11,17–19] aimed at overcoming the intrinsic limit of subjective visually-performed tests.

Among the most used, the SNR-based evaluation at several depths deserves to be mentioned [20] since it is reported as a standard reference method in [17]. In particular, such methods involve the storage of images captured from a US phantom [7,19,21] and consequently DOP is evaluated in the post-processing phase as the maximum depth at which the estimated SNR exceeds a set threshold. Despite the development of such methods, it should be pointed out that the most accredited ones are based on a threshold value subjectively chosen by the operator [4] or without clarifying the choice criteria [16,19]. This constitutes a source of error and unreliability because such methods suffer from an indirect operator subjectivity.

The contribution of the present work lies firstly in the fact that the proposed method is no longer based on threshold values chosen by users or operators but it determines the thresholds in an a totally independent and automatic manner—this is a great advantage in terms of reproducibility and reliability of the results. Secondly, in order to validate this novel method, it was compared with another operator-independent algorithm, already proposed in literature by the same authors [18], whose complexity may make it harder to employ. Therefore, the goal of this study was to improve and validate the functionality of the above method, based also on a previous preliminary study [22], the adaptive SNR threshold method (AdSTM), and compare the measurement results with both the tangent threshold method (TTM) in [18] and the mean judgment of eight different observers. In this study, the AdSTM was tested on four US scanners, each one equipped with three different probe models. Finally, many Monte Carlo simulations (MCS) were carried out in order to quantify the method uncertainty in the DOP measurement.

2. Materials and Methods

The AdSTM is a SNR-based method developed to assess the maximum depth of penetration on the basis of phantom and in-air clips obtained with the transducer on the phantom and held in-air, respectively. The data acquired was processed in MATLAB, obtaining a single image for both phantom and in-air clips (I and I_a), which were the result of the average of the first 15 frames. Then, the proper diagnostic average images (I_c and $I_{c,a}$) were obtained after cropping through automatic masking of I and I_a to remove all the US settings details. In order to compute the SNR as a function of depth, the signal $S(z)$ and the noise $N(z)$ were estimated according to the mean pixel value on a row within a

rectangular region of interest (ROI), automatically selected on I_c and $I_{c,a}$, respectively. An accurate estimation of DOP is closely related to a suitable number of pixels to include in the ROI at each depth. Nevertheless, the main limitation on the number of pixels per depth is due to the phantom section width. Furthermore, other aspects should be taken into account—artifacts at the transducer–phantom (or transducer–air) interface; the possible presence of strong echoes, corresponding to the phantom bottom; and the delimitation of the different attenuation zones. In addition, an excessive pixel number per depth could increase DOP computational cost. Therefore, the minimal pixel number choice (e.g., 30 pixels) is based on a trade-off between an accurate estimation and a limited computational time (about 10 ms). Then, the AdSTM evaluates the corresponding $SNR(z)$ from the ratio $S(z)/N(z)$ and computes the adaptive SNR threshold (SNR_{th}), automatically estimated as:

$$SNR_{th} = S_{max} \cdot \frac{\Delta g_{min}}{L_{fs}}, \tag{1}$$

where S_{max} is the maximum system sensitivity retrieved from the logistic function of the SNR curve [22], Δg_{min} is the minimum gray level difference distinguishable from the human eye [23–25] and L_{fs} is the luminance full scale expressed in terms of gray levels. In particular, the ratio $\Delta g_{min}/L_{fs}$ is always between 0 and 1, while S_{max} was determined according to the following steps:

1. The non-linear curve fitting $f(z)$ was derived, from $SNR(z)$, by the iterative computation of the coefficients $\beta, \chi, \gamma, \eta$ according to the sigmoidal function:

$$f(z) = \frac{\beta}{1 + e^{-\chi z + \gamma}} - \eta, \tag{2}$$

2. The first order derivative $f'(z)$ was calculated and its minimum value z_{min} was used to estimate S_{max} as follows:

$$S_{max} = f(z_{min}), \tag{3}$$

Finally, the AdSTM retrieved the depth value corresponding to the maximum depth of penetration from the intersection between the threshold SNR_{th} and the SNR curve.

In order to test the AdSTM on different US scanners equipped with different probes, the results were compared with the ones obtained from the TTM, an alternative algorithm developed in [18], and the mean judgement of eight independent observers without medical expertise through an eye-based test of I_c (naked eye test). Medical expertise is not mandatory for quality controls of the US systems by means of ad hoc tissue-mimicking phantoms, as they are usually performed by technicians. The protocol employed in this study for the naked eye test was designed by the authors, on the basis of the American Association of Physicists in Medicine (AAPM) report [2]. Each observer independently performed the test making sure to preserve the same test conditions—a well-lit room, maximum monitor light without glare, and keeping a fixed distance from it. Moreover, through an in-house MATLAB software the observer could indicate with a simple mouse click, his decision on I_c . In order to test subjects' inter- and intra-variability, the test was repeated six times providing a randomized image order to the observer, automatically determined by the software.

TTM is an automatic method developed for DOP measurement and based on the gray scale mapping function (GSMF) determination [18] without a threshold subjectively selected by the technician. In particular, after the ROI selection in the diagnostic image, the gray levels were computed as a function of depth, averaging adjacent columns into the ROI. The mean depth profile $g(z)$ obtained, was then interpolated with a non-linear curve, according to the following logistic function (ζ, ν and ξ are constant):

$$b(z) = \frac{\zeta}{1 + e^{-\nu z + \xi}}, \tag{4}$$

Following Equation (4), the DOP was evaluated as the depth at which $b(z)$ intersects the tangent threshold T_{th} , estimated on $b(z)$ as follows:

$$T_{th} = K_{max} \cdot \frac{\Delta g_{min}}{L_{fs}}, \quad (5)$$

where K_{max} is assumed as the maximum system sensitivity and depends both on some characteristics of the medium (i.e., the attenuation constant and a factor related to the absorption in the medium) and some parameters related to US scanner settings (i.e., the bandwidth central frequency, the electronic noise amplitude, and the echo signal amplitude at $z = 0$) as reported in [18]. The main advantage of the AdSTM, as for the TTM, is the automatic determination of the adaptive threshold based on the human eye sensitivity, as defined in Equations (1) and (5), respectively. The arbitrary threshold from the operator is no longer required. On the other hand, the main disadvantage of the TTM is the slow GSMF determination and the high acquisition time.

As in [22], the phantom clips were acquired on a multi-purpose, multi-tissue US phantom (CIRS, Model 040GSE). The device [26] is embedded with test objects of known characteristics and it was designed to provide tissue-mimicking properties under B-mode ultrasound. Such phantoms provide the possibility of testing the US probe on two different zones with attenuation coefficients of 0.70 and 0.50 dB·cm⁻¹·MHz⁻¹. The main characteristics of CIRS 040GSE are reported in Table 1.

Table 1. US phantom characteristics.

Parameter	Characteristics
US phantom model	CIRS 040GSE
Scanning material	Zerdine [®] tissue-mimicking gel
Sound speed (m·s ⁻¹)	1540
Attenuation (dB·cm ⁻¹ ·MHz ⁻¹)	0.70
Dimensions	17.8 × 12.7 × 20.3 cm
Weight	4752 g

Four real-time diagnostic US systems for general purpose imaging (intermediate technology level), indicated as A, B, C, and D, each one of them equipped with three US probes (linear, convex, and vector array, respectively), were tested by means of the phantom zone with 0.70 dB·cm⁻¹·MHz⁻¹ of attenuation, paying attention to display the speckle background only. All the probes were set at their own central frequency among the suitable range (nominal frequency). Moreover, scanning settings (e.g., overall gain, sensitivity time control (STC), and dynamic range) were preserved throughout the clip acquisitions because of their influence on SNR and DOP. In order to compare the different diagnostic systems tested in this work, the phantom and in-air clip acquisitions were carried out according to two different configuration settings (sets 1 and 2 in Table 2), which corresponded to the raw scanner working condition (QA protocols to provide a comparison of the outcomes from different scanners at similar working conditions) and the default preset working condition (clinical preset) provided by the US specialist, respectively. In addition, the ROI was always automatically selected in the same area of the diagnostic image according to the probe model. The influence of the ROI selection was evaluated, in terms of repeatability uncertainty, with changes in ROI size and position by means of Monte Carlo simulations (see the Monte Carlo simulation section). Moreover, a tissue-mimicking phantom with uniform speckle was used in our tests to reduce the aforementioned error component as much as possible.

Table 2. Ultrasound system settings.

Parameter	Settings		
	Set 1	Set 2	
Nominal frequency (MHz)	L: 5-10	C: 2-4	V: 2-3
Dynamic range (dB)	Maximum	Medium	
Field of view (mm)	L: 70 (74 ¹)	C: 180 (184 ¹)	V: 180 (191 ¹)
Focus depth (mm)	L: 20–40	C: 80–120	V: 80–100
Overall gain (dB)	Medium		
STC (dB)	Slide switches aligned in central position ²		
Transmitted power	Maximum	Medium/Maximum	
Zoom	No		
Post-processing	Linear		
Image format	DICOM		

L = linear array probe; C = convex array probe, V = vector/phased array probe. ¹ The field of view (FOV) values are referred to the D diagnostic system only. ² For C probe in set 1 configuration for the D diagnostic system only, STC slide switches have been aligned in right position corresponding to the highest gain level. The US systems in the tests are the following: HS 60 (Samsung), Affiniti 70 (Philips), Acuson S2000 (Siemens), and MyLab X8 (Esaote).

Finally, the compatibility between the different DOP outcomes retrieved from the AdSTM, the TTM, and the naked eye test, were verified according to the following condition [27]:

$$|\mu_1 - \mu_2| \leq \sigma_1 + \sigma_2, \quad (6)$$

where μ_1 and μ_2 are the mean DOP values while σ_1 and σ_2 are the corresponding total standard deviations. If Equation (6) stands, there is no significant discrepancy between the measurements.

3. Monte Carlo Simulation

Monte Carlo simulations (MCSs) can be used to evaluate measurement uncertainty and method robustness in several applications [28–33]; the approach with MCS is also suitable for the uncertainty estimation in DOP measurements as image analysis-based methods [31–33] may not be expressed in an analytical form that is suitable for error analysis by the propagation error law.

As in [22], a first series of MCSs (10^4 cycles) were carried out to evaluate the standard deviation σ (SD) of the SNR threshold (th_{SNR}), through 2.5 and 97.5 percentiles. The following uniform distributions were assigned to the variables influencing th_{SNR} : ROI width l and ROI shift s (s is evaluated with respect to the ROI initial position), both expressed as number of pixels (Table 3a,b). The aforementioned distributions were used, in an analogous way, in order to retrieve the GSMF coefficients contained in the logistic function $b(z)$. Afterwards, a second series of MCSs (10^5 cycles) was carried out for the AdSTM to evaluate DOP distribution histogram for each probe, according to the corresponding th_{SNR} distribution reported in Table 3a,b, together with the previously-assigned distributions l (mean value: 30 px, SD: 1 px) and s , (mean value: 0 px, SD: 2 px). A similar procedure was applied for the TTM where, in addition to l and s distributions, the previously-retrieved GSMF coefficients distributions for DOP assessment were used. The lower number of iterations in the first series was due to the high computational cost of the non-linear curve fitting.

Table 3. (a) Variable settings in Monte Carlo simulations (MCSs) for measurement uncertainty estimation of the AdSTM outcomes for A and B US systems. (b) Variable settings in MCSs for measurement uncertainty estimation of the AdSTM outcomes for C and D US systems.

(a)

US System	Probe	Symbol	Distribution	Mean Value	Standard Deviation	
All	All	l	Uniform	30 px	1 px	
		s	Uniform	0 px	2 px	
A	L	set 1	Uniform	1.91	0.04	
		set 2	Triangular	2.13	0.01	
	C	set 1	Uniform	1.52	0.04	
		set 2	Uniform	0.90	0.04	
	V	set 1	Triangular	1.10	0.02	
		set 2	Triangular	1.63	0.03	
	B	L	set 1	Triangular	1.38	0.03
			set 2	Uniform	1.86	0.06
C		set 1	Triangular	0.49	0.04	
		set 2	Triangular	0.48	0.02	
V		set 1	Triangular	1.76	0.06	
		set 2	Triangular	1.73	0.04	

L = linear array probe; C = convex array probe, V = vector/phased array probe. Mean and SD th_{SNR} values are different for each probe because of the Signal to Noise Ratio (SNR) threshold probe dependency.

(b)

US System	Probe	Symbol	Distribution	Mean Value	Standard Deviation	
All	All	l	Uniform	30 px	1 px	
		s	Uniform	0 px	2 px	
C	L	set 1	Triangular	3.7	0.1	
		set 2	Triangular	12.50	0.01	
	C	set 1	Triangular	2.26	0.02	
		set 2	Triangular	2.78	0.02	
	V	set 1	Triangular	1.94	0.04	
		set 2	Triangular	2.18	0.04	
	D	L	set 1	Uniform	0.54	0.02
			set 2	Triangular	0.74	0.01
C		set 1	Triangular	0.46	0.01	
		set 2	Triangular	1.76	0.01	
V		set 1	Uniform	1.55	0.08	
		set 2	Triangular	2.88	0.01	

L = linear array probe; C = convex array probe, V = vector/phased array probe. Mean and SD th_{SNR} values are different for each probe because of the SNR threshold probe dependency.

4. Results

DOP outcomes for the A, B, C, and D diagnostic systems equipped with the three US probes tested in two configuration settings are reported in Table 4 and shown in Figures 1 and 2. Both the AdSTM

and the TTM repeatability uncertainties were evaluated through 2.5 and 97.5 percentiles of the data distributions from the second series of the aforementioned MCSs. Then, the total uncertainties reported in Table 4 were obtained by combining the above repeatability uncertainties with the probe-dependent uncertainty due to the influence of probe placement on the phantom scanning area, according to [28]. On the other hand, the one-way ANOVA statistical test ($p < 0.001$) was used to test intra- and inter-observer variability in the naked eye test. Globally, by considering the same probe model, independently from the US scanner (except for D scanner), the measurement outcomes are compatible for both the SNR threshold and tangent method and the observers' judgment (Table 4). Nevertheless, the fact that DOP values provided by the AdSTM were generally higher than the TTM ones seems to suggest that the former may be the most sensitive. This may have been due to the different reference profile used in the two algorithms, SNR and gray level, respectively. Such issues seems to be confirmed by the outcomes retrieved for the D scanner. In fact for the latter, the TTM provided incompatible results with the mean observers' judgment, whose DOP values, in turn, almost agreed with the AdSTM outcomes.

Table 4. DOP outcomes for different US diagnostic systems.

US System	Probe	AdSTM (mm)		TTM [18] (mm)		Observers (mm)		
		$\mu \pm \sigma$	%FOV	$\mu \pm \sigma$	%FOV	$\mu \pm \sigma$	%FOV	
A	L	set 1	36 ± 3	4.3	35 ± 2	2.9	39 ± 3	4.3
		set 2	38 ± 2	2.9	38 ± 3	4.3	37 ± 3	4.3
	C	set 1	152 ± 3	1.7	154 ± 6	3.3	138 ± 12	6.7
		set 2	141 ± 2	1.1	124 ± 2	1.1	128 ± 6	3.3
	V	set 1	101 ± 5	2.8	105 ± 4	2.2	107 ± 10	5.6
		set 2	106 ± 3	1.7	86 ± 2	1.1	97 ± 7	3.9
B	L	set 1	43 ± 2	2.9	40 ± 3	4.3	40 ± 3	4.3
		set 2	42 ± 2	2.9	40 ± 2	2.9	40 ± 2	2.9
	C	set 1	99 ± 2	1.1	84 ± 3	1.7	83 ± 6	3.3
		set 2	89 ± 2	1.1	78 ± 3	1.7	77 ± 7	3.9
	V	set 1	120 ± 2	1.1	113 ± 5	2.8	109 ± 8	4.4
		set 2	102 ± 2	1.1	100 ± 4	2.2	98 ± 10	5.6
C	L	set 1	46 ± 2	2.9	39 ± 2	2.9	41 ± 4	5.7
		set 2	48 ± 2	2.9	47 ± 2	2.9	48 ± 3	4.3
	C	set 1	151 ± 6	3.3	142 ± 5	2.8	143 ± 8	4.4
		set 2	140 ± 2	1.1	142 ± 6	3.3	135 ± 5	2.8
	V	set 1	124 ± 2	1.1	99 ± 2	1.1	118 ± 14	7.8
		set 2	118 ± 9	5.0	120 ± 5	2.8	112 ± 13	7.2
D	L	set 1	43 ± 2	2.7	28 ± 2	2.7	37 ± 4	5.4
		set 2	40 ± 2	2.7	28 ± 2	2.7	35 ± 2	2.7
	C	set 1	107 ± 2	1.1	−1	−1	81 ± 8	4.3
		set 2	104 ± 2	1.1	61 ± 3	1.6	96 ± 6	3.3
	V	set 1	138 ± 6	3.1	78 ± 3	1.6	130 ± 7	3.7
		set 2	141 ± 2	1.0	86 ± 2	1.0	138 ± 7	3.7

L = linear array probe; C = convex array probe, V = vector/phased array probe. Results are reported in terms of mean ± SD ($\mu \pm \sigma$). The percentage error is expressed with respect to the field of view ($\%FOV = 100 (\sigma / FOV)$) whose values are reported in Table 2. ¹ Missing outcome.

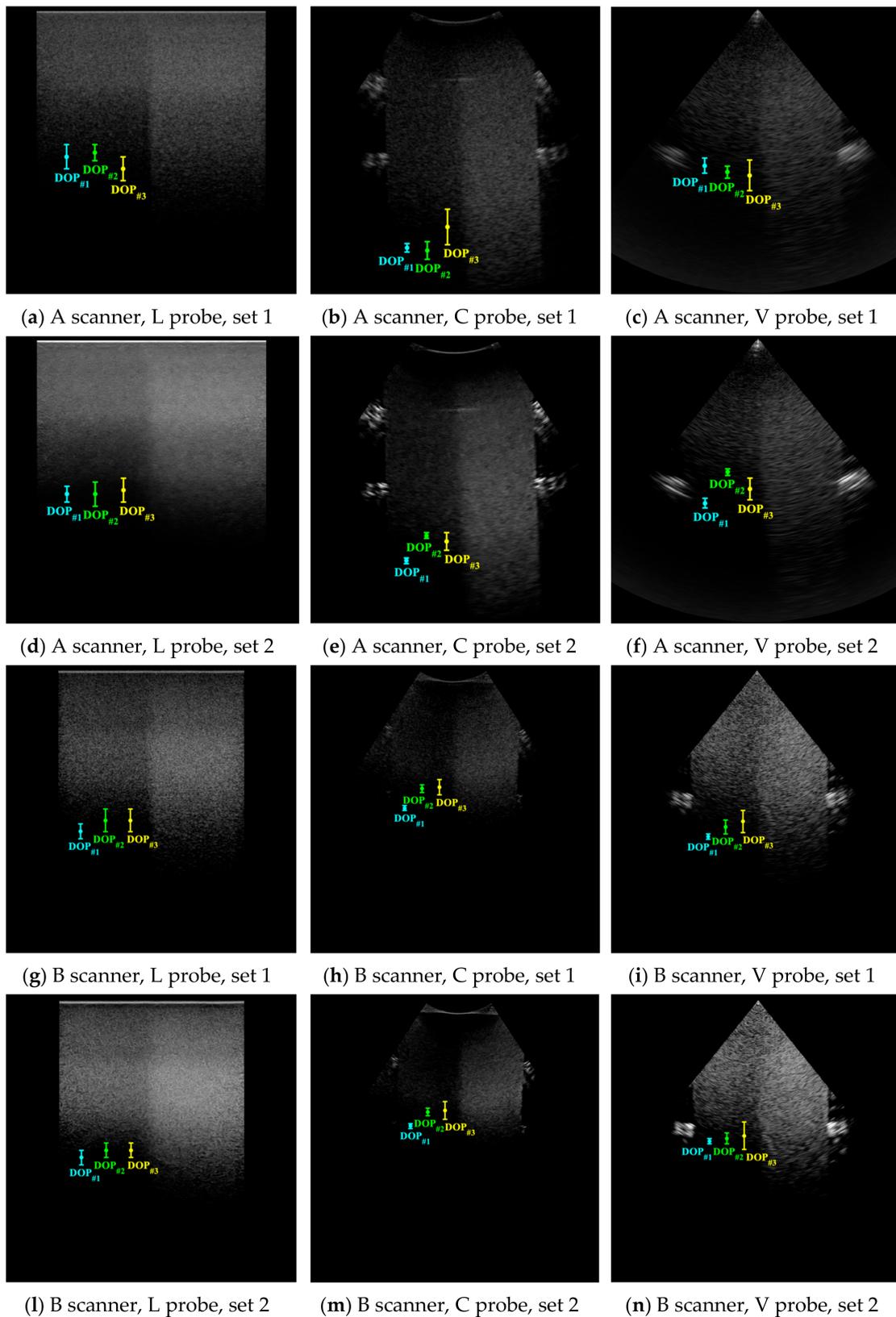


Figure 1. Depth of penetration for A and B diagnostic systems equipped with linear L, convex C, vector V probes tested in sets 1 and 2. DOP_{#1} and DOP_{#2} were evaluated through the AdSTM and the TTM [18], respectively, while DOP_{#3} was assessed by naked eye. DOP values are represented with related error bars due to the total uncertainties reported in Table 4.

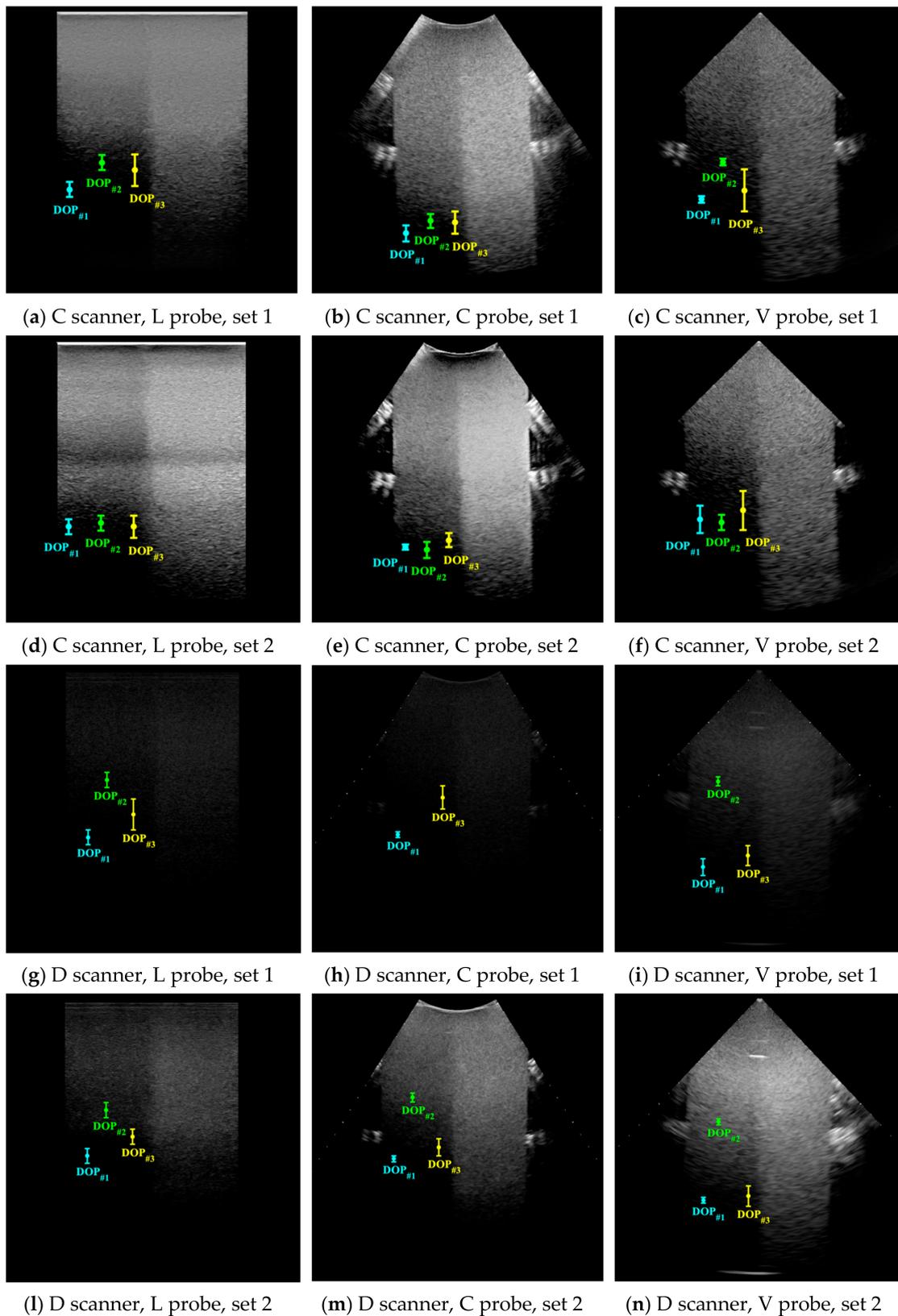


Figure 2. Depth of penetration for C and D diagnostic systems equipped with linear L, convex C, vector V probes tested in sets 1 and 2. DOP_{#1} and DOP_{#2} have been evaluated through the AdSTM and the TTM [18], respectively, while DOP_{#3} was assessed by naked eye. DOP values are represented with related error bars due to the total uncertainties reported in Table 4.

Furthermore, the only missing outcome reported in Table 4 was due to the TTM impossibility of retrieving a DOP value for the convex probe in set 1 configuration, despite the highest STC setting. A possible cause may be found in the intrinsic low level of the D scanner dynamic; such a limit is mainly emphasized in correspondence of the raw scanner working conditions (set 1) resulting in very darker US images, as shown in Figure 2g–i.

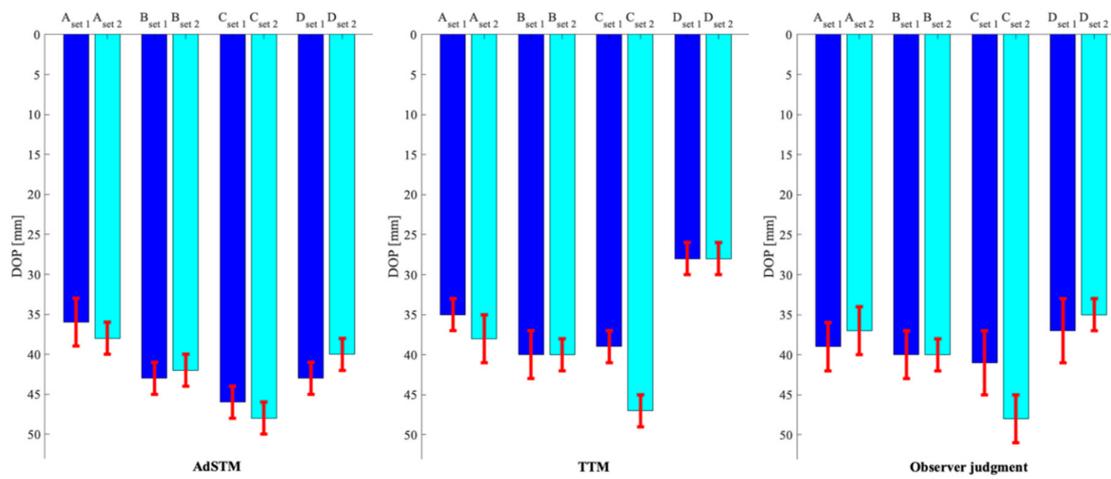
5. Discussion

By focusing attention on the linear array probe for the first three scanners, the outcomes of both the automatic methods were mostly compatible with the mean observers' judgment. Moreover, there was not a significant discrepancy between DOP values in the two configuration settings. On the other hand, the results retrieved for the convex array probe showed a slight disagreement between the two automatic methods, despite the agreement between the TTM and the scores provided by the observers. Nevertheless, it can be seen that mean DOP values, assessed with both automatic methods, were usually higher in the set 1 configuration than the corresponding ones in set 2 for all the US scanners. This phenomenon may suggest that in the default preset condition, the scanner image pre- and post-processing seemed to improve the image quality to the detriment of the maximum penetration depth. For the vector array probe both automatic methods were generally compatible with the visually-assessed DOP values (excluding D scanner results). Lastly, a further consideration is that, independently from the scanner considered, among the US probes, the linear array showed the higher compatibility occurrence for the results obtained with the three different protocols. Furthermore, it can be noticed that such occurrence is mainly verified for C scanner.

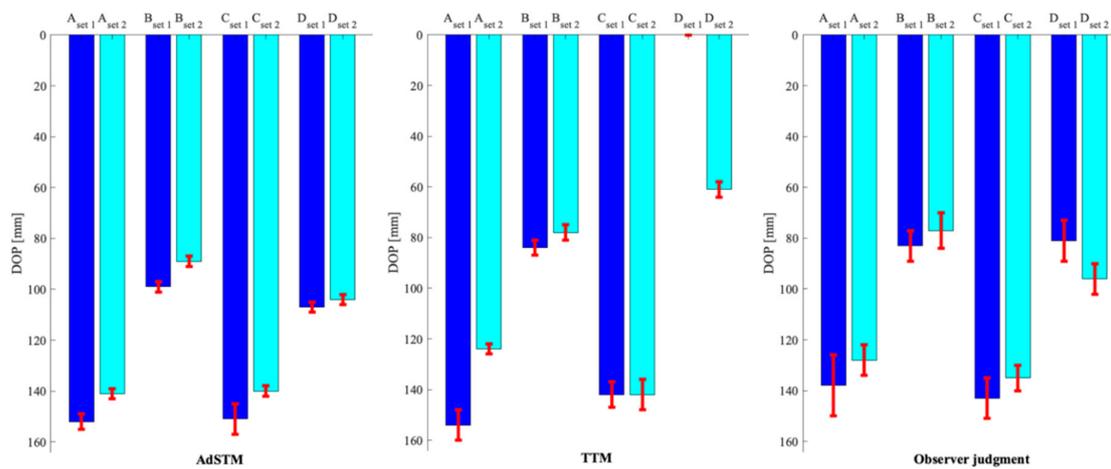
As regards the percentage error for the AdSTM and the TTM, for A and B scanners, the highest and the lowest percentage errors were observed for the linear array (4.3%) and for the convex array (1.1%) probes, respectively. On the other hand, for C and D scanners, considering the AdSTM only, the highest (5.0% and 3.1%) and the lowest (1.1% and 1.0%) percentage errors were always observed for the vector array probe, whereas, by considering the TTM for C scanner only, the highest and the lowest percentage errors were observed for the convex array (3.3%) and for the vector array (1.1%) probes. Furthermore, considering A, B, and C scanners, data confirm the equivalence between the two automatic methods in evaluating DOP uncertainty. In fact, the mean percentage error for the AdSTM was around 2.3%, whereas for the TTM was around 2.6%. Conversely, the mean percentage error for the visual assessment was around 4.5%, therefore confirming the higher dispersion of the naked eye method in DOP assessment with respect to the automatic methods. Nevertheless, it should be noticed that the lowest observers' absolute uncertainty was obtained for the linear array probe, which was likely due to the differences in probe technology and scanning format between the linear and vector/convex array models.

Finally, the performances of the four US systems for the AdSTM, the TTM, and naked eye test were compared on the basis of the highest and lowest DOP mean values. For the linear array probe models (Figure 3a) the best performance was provided by the C scanner (set 2) for all the methods, while the worst performance was given by the A scanner (set 1) for the AdSTM and the D scanner (set 2) for the TTM and naked eye method. On the other hand, among the convex array probe models (Figure 3b), the highest DOP mean value was achieved for the A scanner (set 1) applying the automatic methods, whereas according to the observer judgement such value was obtained by the C scanner (set 1). Besides, the lowest DOP mean value could be found for the B scanner (set 2) in the AdSTM and naked eye method, while for the D scanner (set 2) it was the TTM. Moreover, for vector array probe models (Figure 3c) the AdSTM and naked eye method agreed in attributing the best performance to the D scanner (set 2), while for the TTM the C scanner (set 2) provided the highest depth of penetration mean value. Conversely, the worst performance was provided by the A scanner in the AdSTM and naked eye method (sets 1 and 2, respectively), while the D scanner (set 1) this was for the TTM. In general, considering the same probe model, it can be noticed that no significant differences in the results presented for the four systems were detected due to changes of the method applied.

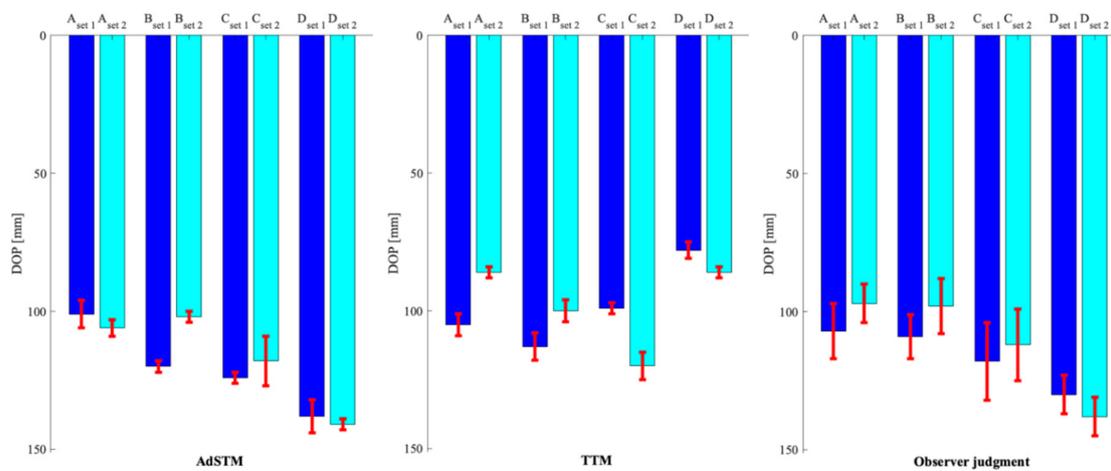
In particular, the highest and the lowest DOP mean values were shown in correspondence of the same US system, except for cases related to the D scanner, mainly due to the intrinsic low level of the D scanner brightness level that prevented the TTM from providing consistent results.



(a) Linear array probes



(b) Convex array probes



(c) Vector array probes

Figure 3. Depth of penetration outcomes retrieved from the AdSTM, the TTM, and observer judgment for (a) linear, (b) convex, and (c) vector array probe models.

However, even if the results are very promising, some limitations should be taken into account for further developments—the influence of US system settings deserves an in-depth investigation and, in addition, more tests on different US phantoms should be carried out.

6. Conclusions

The adaptive SNR threshold method (AdSTM), developed in a previous preliminary study, was tested on a sample of medical US scanners in two different configuration settings. In particular, data was acquired on the same tissue-mimicking phantom from four US diagnostic systems, each equipped with three different US probes that worked at their own central frequency. The AdSTM outcomes were compared with the results from: (a) another automatic method taken from the current state of the art (tangent threshold method, TTM) developed by the same authors and, (b) the mean judgment of eight different observers without medical expertise. Each observer independently performed the test through an in-house MATLAB software preserving the same acquisition conditions. Both AdSTM and TTM repeatability uncertainties were estimated through MCS and subsequently combined with the probe placement uncertainty.

The outcomes usually showed measurement compatibility among both the automatic methods and the observers' judgment. However, it should be noted that the naked eye test scores' compatibility was conditioned by the large observer sample and by the high number of test repetitions. In QA programs where DOP assessment is still visually carried on by a single operator, such compatibility could be no longer guaranteed, whereas, the application of the AdSTM and the TTM assure reproducibility and traceability because of the algorithm threshold that does not depend on the subjectivity of the technician. Nevertheless, the AdSTM seems to have a higher sensitivity than the TTM because of higher DOP values. Both methods have the advantage of the simplicity in electronic noise estimation because, in the first case it can be directly retrieved from the in-air clip, while in the second, such problems are bypassed through the GSMF evaluation. Nevertheless, in correspondence of a US scanner with limited gray level dynamic range, the TTM seemed to be less reliable. Moreover, the AdSTM required lower computing performance and lower acquisition as well as processing time, excluding the (same) probe positioning and data transfer procedure. With regard to the acquisition time, the TTM needed the acquisition of at least 60 images by varying the STC configuration and the overall gain for the GSMF evaluation, whereas in the AdSTM the acquisition of only two clips of 2 s each one with a US scanner frame rate in the range of 20-30 Hz was required. As a final remark, it should be pointed out that the AdSTM application usually took half the time of the TTM. On the basis of the above results and their limitations, further developments of the present study could be the application of the AdSTM to DOP measurements on (a) a wide number of US diagnostic systems available in the market and (b) different US phantom models, as well as (c) changing US system settings. Such follow-up studies will allow for the selection of the best US system settings for DOP measurement and the improvement of AdSTM robustness and will contribute to providing a reference measurement standard in ultrasound DOP measurements.

Author Contributions: Conceptualization, G.F. and A.S.; methodology, G.F. and F.F.; software, G.F.; validation, G.F., F.F., and A.S.; formal analysis, G.F.; investigation, G.F., F.F., and A.S.; resources, J.G. and S.A.S.; data curation, G.F.; writing—original draft preparation, G.F. and F.F.; writing—review and editing, G.F., F.F., A.S., J.G., S.C., and S.A.S.; visualization, G.F.; supervision, A.S., S.C., and S.A.S.; project administration, A.S.; funding acquisition, J.G. and S.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors wish to thank SAMSUNG Healthcare, PHILIPS Healthcare, SIEMENS Healthineers, and ESAOTE for hardware supply and technical assistance in data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lu, Z.F.; Hangiandreou, N.J.; Carson, P. Clinical ultrasonography physics: state of practice. In *Clinical Imaging Physics: Current and Emerging Practice*, 1st ed.; Samei, E., Pfeiffer, D.E., Eds.; Wiley Blackwell: Hoboken, NJ, USA, 2020; pp. 261–286.
2. Goodsitt, M.M.; Carson, P.L.; Witt, S.; Hykes, D.L.; Kofler, J.M., Jr. Real-time B-mode ultrasound quality control test procedures, Report of AAPM Ultrasound Task Group No. 1. *Med. Phys.* **1998**, *25*, 1385–1406. [[CrossRef](#)] [[PubMed](#)]
3. Scorza, A.; Orsini, F.; De Marchis, C.; Caramia, C.; Sciuto, S.A.; Galo, J. A Comparative study on the influence of phantoms and test objects on quality control measurements in B-mode ultrasound systems: Preliminary results. In Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy, 11–13 June 2018. [[CrossRef](#)]
4. Gibson, N.M.; Dudley, N.J.; Griffith, K. A computerised quality control testing system for B-mode ultrasound. *Ultrasound Med. Biol.* **2001**, *27*, 1697–1711. [[CrossRef](#)]
5. Thijssen, J.M.; Weijers, G.; de Korte, C.L. Objective performance testing and quality assurance of medical ultrasound equipment. *Ultrasound Med. Biol.* **2007**, *33*, 460–471. [[CrossRef](#)] [[PubMed](#)]
6. Hangiandreou, N.J.; Stekel, S.F.; Tradup, D.J.; Gorny, K.R.; King, D.M. Four-year experience with a clinical ultrasound quality control program. *Ultrasound Med. Biol.* **2011**, *37*, 1350–1357. [[CrossRef](#)] [[PubMed](#)]
7. Tradup, D.J.; Hangiandreou, N.J.; Taubel, J.P. Comparison of ultrasound quality assurance phantom measurements from matched and mixed scanner-transducer combinations. *J. Appl. Clin. Med. Phys.* **2003**, *4*, 239–247. [[CrossRef](#)] [[PubMed](#)]
8. Goldman, L.W.; Fowlkes, B.J. *Medical CT & Ultrasound: Current Technology & Applications—AAPM Summer School 1995*; American Institute of Physics: College Park, MD, USA, 1995.
9. Sassaroli, E.; Crake, C.; Scorza, A.; Kim, D.-S.; Park, M.-A. Image quality evaluation of ultrasound imaging systems: Advanced B-modes. *J. Appl. Clin. Med. Phys.* **2019**, *20*, 115–124. [[CrossRef](#)] [[PubMed](#)]
10. American institute of Ultrasound in Medicine & AIUM Standards Committee. *Standard Methods for Measuring Performance of Pulse-Echo Ultrasound Equipment*; The Institute: Rockville, MD, USA, 1991.
11. AIUM Technical Standards Committee, Quality Assurance Subcommittee. *AIUM Quality Assurance Manual for Gray-Scale Ultrasound Scanners*; American institute of Ultrasound in Medicine: Laurel, MD, USA, 1995.
12. Browne, J.E.; Watson, A.J.; Muir, C. An Investigation of the Relationship between In-vitro and In-vivo Ultrasound Image Quality Parameters. *Ultrasound* **2004**, *12*, 202–210. [[CrossRef](#)]
13. Cozzolino, P.; Stramare, R.; Udilano, A.; Castoro, M.; Scagliori, E.; Beltrame, V.; Rubaltelli, L. Quality control of ultrasound transducers: Analysis of evaluation parameters and results of a survey of 116 transducers in a single hospital. *Radiol. Med.* **2010**, *115*, 668–677. [[CrossRef](#)] [[PubMed](#)]
14. Fabiszewska, E.; Pasicz, K.; Grabska, I.; Skrzyński, W.; Ślusarczyk-Kacprzyk, W.; Bulski, W. Evaluation of Imaging Parameters of Ultrasound Scanners: Baseline for Future Testing. *Pol. J. Radiol.* **2017**, *82*, 773–782. [[CrossRef](#)] [[PubMed](#)]
15. Mannila, V.; Sipilä, O. Phantom-based quality assurance measurements in B-mode ultrasound. *Acta Radiol. Short Rep.* **2013**, *2*, 1–4. [[CrossRef](#)] [[PubMed](#)]
16. Dudley, N.J.; Gibson, N.M. Early experience with automated B-mode quality assurance tests. *Ultrasound* **2014**, *22*, 15–20. [[CrossRef](#)] [[PubMed](#)]
17. Ultrasonics—Pulse-Echo Scanners—Part 2: Measurement of Maximum Depth of Penetration and Local Dynamic Range, 1st ed. International Standard IEC 61391-2: Geneva, Switzerland, 2010.
18. Scorza, A.; Lupi, G.; Sciuto, S.A.; Bini, F.; Marinozzi, F. A novel approach to a phantom based method for maximum depth of penetration measurement in diagnostic ultrasound: A preliminary study. In Proceedings of the 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Turin, Italy, 7–9 May 2015. [[CrossRef](#)]
19. Gorny, K.R.; Tradup, D.J.; Hangiandreou, N.J. Implementation and validation of three automated methods for measuring ultrasound maximum depth of penetration: Application to ultrasound quality control. *Med. Phys.* **2005**, *32*, 2615–2628. [[CrossRef](#)] [[PubMed](#)]
20. Sassaroli, E.; Scorza, A.; Crake, C.; Sciuto, S.A.; Park, M.-A. Breast Ultrasound Technology and Performance Evaluation of Ultrasound Equipment: B-Mode. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2017**, *64*, 192–205. [[CrossRef](#)] [[PubMed](#)]

21. Kollmann, C.; deKorte, C.; Dudley, N.J.; Gritzmann, N.; Martin, K.; Evans, D.H. Guideline for Technical Quality Assurance (TQA) of ultrasound devices (B-mode)—Version 1.0 (July 2012): EFSUMB Technical Quality Assurance group—US-TQA/B. *Ultraschall Med.* **2012**, *33*, 544–599. [[CrossRef](#)] [[PubMed](#)]
22. Fiori, G.; Fuiano, F.; Vurchio, F.; Scorza, A.; Schmid, M.; Conforto, S.; Sciuto, S.A. A preliminary study on a novel method for depth of penetration measurement in ultrasound quality assessment. In Proceedings of the 24th IMEKO TC4 International Symposium & 22nd International Workshop on ADC Modelling and Testing, Palermo, Italy, 14–16 September 2020.
23. Kimpe, T.; Tuytschaever, T. Increasing the number of gray shades in medical display systems—How much is enough? *J. Digit. Imaging* **2007**, *20*, 422–432. [[CrossRef](#)] [[PubMed](#)]
24. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2002; pp. 35–98.
25. Gille, J.; Arend, L.; Larimer, J.O. Display characterization by eye: Contrast ratio and discrimination throughout the grayscale. In Proceedings of the SPIE 5292, Human Vision and Electronic Imaging IX, San Jose, CA, USA, 7 June 2004. [[CrossRef](#)]
26. CIRS Tissue Simulation & Phantom Technology. Multi-Purpose Multi-Tissue Ultrasound Phantom—Model 040GSE. Available online: <http://www.cirsinc.com/wp-content/uploads/2020/03/040GSE-DS-032320-1.pdf> (accessed on 10 September 2020).
27. Taylor, J.R. An. *Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books: Sausalito, CA, USA, 1996; pp. 13–44.
28. Scorza, A.; Conforto, S.; D’Anna, C.; Sciuto, S.A. A Comparative Study of the Influence of Probe Placement on Quality Assurance Measurements in B-mode Ultrasound by Means of Ultrasound Phantoms. *Open Biomed. Eng. J.* **2015**, *9*, 164–178. [[CrossRef](#)] [[PubMed](#)]
29. Fiori, G.; Fuiano, F.; Scorza, A.; Schmid, M.; Conforto, S.; Sciuto, S.A. ECG waveforms reconstruction based on equivalent time sampling. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1–3 June 2020. [[CrossRef](#)]
30. Orsini, F.; Fuiano, F.; Fiori, G.; Scorza, A.; Sciuto, S.A. Temperature influence on viscosity measurements in a rheometer prototype for medical applications: A case study. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019. [[CrossRef](#)]
31. Orsini, F.; Vurchio, F.; Scorza, A.; Crescenzi, R.; Sciuto, S.A. An Image Analysis Approach to Microgrippers Displacement Measurement and Testing. *Actuators* **2018**, *7*, 64. [[CrossRef](#)]
32. Vurchio, F.; Orsini, F.; Scorza, A.; Fuiano, F.; Sciuto, S.A. A preliminary study on a novel automatic method for angular displacement measurements in microgripper for biomedical applications. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1–3 June 2020. [[CrossRef](#)]
33. Fiori, G.; Fuiano, F.; Scorza, A.; Galo, J.; Conforto, S.; Sciuto, S.A. Lowest detectable signal in medical PW doppler quality control by means of a commercial flow phantom: A case study. In Proceedings of the 24th IMEKO TC4 International Symposium & 22nd International Workshop on ADC Modelling and Testing, Palermo, Italy, 14–16 September 2020.

