A Very Large Cardiac Channel Data Database (VLCD) used to Evaluate Global Image Coherence (GIC) as an In-Vivo Image Quality Metric

Ole Marius Hoel Rindal, Member, IEEE, Tore Grüner Bjåstad, Member, IEEE, Torvald Espeland, Erik Andreas Rye Berg, and Svein Erik Måsøy

Abstract— Ultrasound image quality is of utmost importance for a clinician to reach a correct diagnosis. Conventionally, image quality is evaluated using metrics to determine the contrast and resolution. These metrics requires localization of specific regions and targets in the image such as a region of interest (ROI), a background region, and or, a point scatterer. Such objects can all be difficult to identify in in-vivo images, especially for automatic evaluation of image quality in large amounts of data. Using a matrix array probe, we have recorded a Very Large cardiac Channel data Database (VLCD) to evaluate coherence as an in-vivo image quality metric. The VLCD consists of 33 280 individual image frames from 538 recordings of 106 patients. We also introduce a Global Image Coherence (GIC), an in-vivo image quality metric that does not require any identified ROI since it is defined as an average coherence value calculated from all the data pixels used to form the image, below a pre-selected range. The GIC is shown to be a quantitative metric for in-vivo image quality when applied to the VLCD. We demonstrate, on a subset of the dataset, that the GIC correlates well with the conventional metrics contrast ratio (CR) and the generalized contrast-to-noise ratio (gCNR) with R=0.74 ($p<0.005$) and R=0.62 ($p<0.005$) respectively. There exists multiple methods to estimate the coherence of the received signal across the ultrasound array. We further show that all coherence measures investigated in this study are highly correlated ($R>0.9$, $p<0.001$) when applied to the VLCD. Thus, even though there are differences in the implementation of coherence measures, all quantify the similarity of the signal across the array and can be averaged into a GIC to evaluate image quality automatically and quantitatively.

Index Terms— Software beamforming, image quality, adaptive beamforming, coherence, matrix array, channel data.

This work was supported by the Center for Innovative Ultrasound Solutions (CIUS) and the Research Council of Norway under Project 237887.

Rindal is affiliated with the Research Group for Digital Signal Processing and Image Analysis, Department of Informatics, University of Oslo, Oslo, Norway (e-mail: omrindal@ifi.uio.no).

Bjåstad is with GE Healthcare, GE Vingmed Ultrasound AS, Horten, Norway.

Espeland, Rye Berg and Måsøy are with the Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

I. Introduction

VARYING image quality in ultrasound cardiac imaging results in an undesirable lottery for the clinician performing the scan. Image quality varies from patient to patient, from view to view and between clinicians. Slight adaptations to the probe's position in relation to the ribs can even change image quality from frame to frame. Despite its variability, image quality is of utmost importance for clinicians to reach the correct diagnosis [1]. In echocardiography image quality is degraded from the common sources of noise in ultrasound imaging; phase-aberration, reverberation clutter, off-axis scattering and thermal noise. A quantitative metric of ultrasound image quality could assist the clinician performing the scan, and allow an automatic selection of the recording with the highest image quality, possibly improving diagnostics. Another important aspect is that such a metric may give an indication on how reliable the measurements made in a certain image are. This is useful both while recording and measuring, and when reviewing a list of measurements made on a recording. Lastly, a quantitative metric can be used to optimize imaging parameters and beamforming methods.

The flexibility of software beamforming has introduced a myriad of adaptive beamforming methods presented in the literature [2]–[7] . Adaptive beamforming methods aim at improving image quality by adapting the processing based on the received signals. A second aspect of software beamforming is the ability to easily collect raw channel data, allowing processing and further analysis of data offline. This results in an unprecedented flexibility for researchers to prototype new beamforming methods on relevant clinical in-vivo data. We have recorded in-vivo cardiac channel data on a GE Vingmed Ultrasound Vivid E95 ultrasound system using the 4Vc-D matrix array probe (GE Vingmed Ultrasound AS, Horten, Norway) in a Very Large Cardiac channel data Database (VLCD) consisting of 33 280 individual image frames from 538 recordings of 106 patients.

One category of adaptive beamformers are often denoted coherence beamformers, since they in various ways utilize the coherence, or similarity, of the received sig-

nals. There are several ways to measure the coherence, presented in the literature. We will briefly review some of the most used coherence measures in Section II. Some of these calculate spatial coherence as the similarity of the delayed signals across array elements as a function of element separation or lag. One of these spatial coherence measures, the Lag-One Coherence (LOC), was suggested as a metric for ultrasound image quality in [8]. Long et. al. argue that LOC and coherence in general is sensitive to all the major forms of ultrasonic noise. Coherence is reduced by focusing errors, phase aberrations and off-axis scattering since the delayed signals across the aperture in these cases are decorrelated [9] [10]. Thermal noise lowers the coherence since uncorrelated noise across the delayed array element signals introduce a delta function in the spatial coherence that scales with amplitude based on the relative noise power [11]. Additionally, reverberation clutter has been shown to have a similar effect on the coherence as thermal noise [12].

Conventionally, image quality of ultrasound images is evaluated by measuring contrast and resolution. The most widely used contrast metrics are the contrast ratio (CR) [13], contrast-to-noise ratio (CNR) [14] and more recently the generalized contrast-to-noise ratio (gCNR) [15]. Even though both CR and CNR are shown to correlate strongly with assessments by human observers, they can both be manipulated by alterations to the dynamic range of images [16]. While the gCNR is immune to dynamic range alterations [15], all three of these contrast metrics have a major drawback for in-vivo usability, namely the need to identify two regions; a region-of-interest (ROI) usually tissue, against a background typically consisting of noise. Such regions can be hard to identify in in-vivo images and often require manual interaction. The LOC can be used to measure in-vivo image quality, since it is a single ROI measurement. However, as described in [8] an ROI needs to be identified.

We will further elaborate on the results from [8], and introduce and demonstrate a metric we denote Global Image Coherence (GIC). GIC does not require any identified or segmented ROI since it is defined as an average coherence value calculated from all the data pixels used to form the image, below a pre-selected range. We will also study the similarity between various available coherence measures. More formally, the vast amount of recorded channel data in the VLCD allows us to empirically test the following hypotheses:

i) published coherence measures are strongly correlated.
ii) Global Image Coherence (GIC) can be used as a quantitative metric for in-vivo image quality.

The current article also builds upon our recent study [17] where we used a variant of the GIC to estimate image quality improvements resulting from an aberration correction algorithm. This was a clinical study where four clinicians evaluated cardiac cineloops with and without aberration correction in a blinded and left-right randomized side-by-side setup. However, in that study GIC was used as a metric on the same channel data, with aberration correction processing as the only difference between the images. In this manuscript we aim to use GIC as a general evaluation of image quality comparing images from different recordings, different patients, and different views.

This article is organized as follows. Section II briefly review some of the most used coherence measures introduced in the literature, which we further investigate and compare. Section III describes the VLCD, and the details of the implemented beamforming, automated cardiac view classification, and statistical analysis used in this article. Here, we also define the Global Image Coherence mathematically. Section IV presents the results of the various coherence measures applied to the entire database, as well as using GIC to evaluate image quality. Results are discussed in Section V with concluding remarks in Section VI.

II. Background and Theory

Let us assume a 2D matrix array of M elements in the azimuth direction (along the x-axis) and N elements in the elevation direction (along the y-axis) in a Cartesian coordinate system. In order to acquire 2D images, in the xz-plane, we transmit focused beams steered in different directions in this plane. A received channel signal on element m in azimuth, and n in elevation, with the appropriate propagation time delay applied for a specific pixel (x, z), is here defined as

$$s_{mn}(x, z) \equiv s_{mn}. \tag{1}$$

The following equations, in sub-section II A to F describe the pixel value in the image b for a given beamforming method. A 2D image is acquired specifying the pixels in the 2D xz plane, however it can be extended to calculate values for a full 3D volume in xyz.

In this article, we focus on 2D sector scan images acquired by a cardiac 2D matrix array transducer. Since these are sector scan images we define the pixels in polar coordinates using an angle θ and a range depth r.

A. Delay-and-sum (DAS)

The conventional DAS implementation is the coherent combination of the signals received by all elements, yielding

$$b_{\mathrm{DAS}} = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{mn} s_{mn}, \tag{2}$$

where w_{mn} is the receive apodization for element mn, a static term often determined from the F-number and pixel depth r.

B. Coherence Factor (CF)

The CF was first introduced by Mallart and Fink in 1994 [9], as the ratio between the coherent and incoherent

energy across the aperture:

$$\mathrm{CF} = \frac{\left|\sum_{m=1}^{M}\sum_{n=1}^{N} s_{mn}\right|^2}{MN\sum_{m=1}^{M}\sum_{n=1}^{N}|s_{mn}|^2}. \tag{3}$$

The CF has been used as an adaptive weight to increase image quality [2] as;

$$b_{\mathrm{CF}} = \mathrm{CF}\, b_{\mathrm{DAS}}. \tag{4}$$

C. Phase Coherence Factor (PCF)

The PCF was introduced by Camacho et al. in 2009 [3] as

$$\mathrm{PCF} = \max\left\{0, 1 - \frac{\gamma}{\sigma_0}p\right\}, \tag{5}$$

where γ is a parameter to adjust the sensitivity of PCF to out-of-focus signals, $\sigma_0 = \pi/\sqrt{3}$ is the nominal standard deviation of a uniform distribution between $-\pi$ and π, and p is given by

$$p = \min\left\{\sigma\left(\boldsymbol{\phi}\right), \sigma\left(\boldsymbol{\phi}^A\right)\right\}, \tag{6}$$

where $\boldsymbol{\phi} = [\phi_{1,1}\ \phi_{1,2}\ldots\phi_{M,N}]$ is the instantaneous phase across the 2D aperture, and $\sigma(\boldsymbol{\phi})$ is its standard deviation. We use $\gamma = 1$ in this study. To avoid phase wrapping discontinuity a set of auxiliary phases $\boldsymbol{\phi}^A = [\phi^A_{1,1}\ \phi^A_{1,2}\ldots\phi^A_{M,N}]$ is computed as

$$\phi^A_{mn} = \begin{cases} \phi_{mn} + \pi & \text{if } \phi_{mn} < 0, \\ \phi_{mn} - \pi & \text{otherwise.} \end{cases} \tag{7}$$

The beamformed image is computed using PCF as an adaptive weight

$$b_{\mathrm{PCF}} = \mathrm{PCF}\, b_{\mathrm{DAS}}. \tag{8}$$

D. Circular Coherence Factor (CCF)

Comacho et al. published in 2011 a slight modification of the PCF, namely the Circular Coherence Factor (CCF) [18]. The CCF is defined as the square root of the variance of the instantaneous phase across the aperture $\boldsymbol{\phi}$,

$$\mathrm{CCF} = 1 - \sqrt{\mathrm{var}(\cos\boldsymbol{\phi}) + \mathrm{var}(\sin\boldsymbol{\phi})}. \tag{9}$$

The CCF is described to "[...] fall off faster than PCF, representing a stricter focusing quality measurement" [18]. The CCF can also be used as an adaptive weight to the beamformed image so that

$$b_{\mathrm{CCF}} = \mathrm{CCF}\, b_{\mathrm{DAS}}. \tag{10}$$

E. Short Lag Spatial Coherence (SLSC)

The short lag spatial coherence (SLSC) algorithm was introduced by Lediju Bell et al. in [4]. Even though SLSC can be calculated for a 2D matrix array [19], we will simplify our implementation and collapse the elevation dimension of the array by summing the N elements in elevation after the propagation delay have been applied. The spatial correlation for a 1D array can be calculated as

$$\hat{R}(l) = \frac{1}{M-l}\sum_{m=1}^{M-l}\frac{\sum_{r=r_1}^{r_2} s_m(r)s_{m+l}(r)}{\sqrt{\sum_{r=r_1}^{r_2} s_m^2(r)s_{m+l}^2(r)}}, \tag{11}$$

where s is the delayed signal, r is the depth sample index, l is the distance, or lag, in number of elements between two points on the aperture. The sum over r results in a correlation over a given kernel size, $r_2 - r_1$ of pixels. The short lag spatial coherence, is calculated as the sum over the first Q lags,

$$b_{\mathrm{SLSC}} = \sum_{l=1}^{Q}\hat{R}(l). \tag{12}$$

Thus, notice that b_{SLSC} is an image of the coherence and not the backscattered signal amplitude as with DAS. The SLSC is a visualization of the spatial coherence of backscattered ultrasound waves, building upon the theoretical prediction of the van Cittert-Zernike (VCZ) theorem. The implications of the VCZ theorem for pulse-echo ultrasonic imaging is discussed in [9], [20]. In this study, we used $Q = 9$ and a kernel size of λ for SLSC.

F. Lag One Coherence (LOC)

Assuming the simplified implementation where we collapse the elevation dimension of the 2D matrix array, the Lag One Coherence (LOC) [8] can be calculated using the same expression for the spacial coherence as SLSC in equation (11), but only evaluated at lag l = 1.

G. Generalized Coherence Factor (GCF)

Again assuming the simplified implementation where we collapse the elevation dimension of the 2D matrix array, the generalized coherence factor is defined as [2]

$$\mathrm{GCF} = \frac{\sum_{n<M_0}|S_n|^2}{\sum_{n=-\frac{M}{2}}^{\frac{M}{2}-1}|S_n|^2}, \tag{13}$$

where S is the M-point Fourier spectra over the aperture of the delayed channel data,

$$S_n = \sum_{m=0}^{M-1} s_m e^{-j2\pi(m-M/2)d\frac{n}{Md}}, \tag{14}$$

where $n \in [-\frac{M}{2}, \frac{M}{2}-1]$ is the spatial frequency index where M is assumed to be even, d is the pitch of the array, and M_0 is an arbitrary constant within $[0, \frac{M}{2}-1]$ that specifies the low spatial frequency region, thus going from $-M_0$ to M_0. Note that if $M_0 = 0$ the GCF simplifies to the CF. In this study we used $M_0 = 4$. We collapsed the elevation dimension of the 2D matrix array to avoid a 2D Fourier spectra of the 2D aperture data and thus having to introduce a significantly changed implementation of the GCF. The beamformed DAS image can then be multiplied with GCF

$$b_{\mathrm{GCF}} = \mathrm{GCF}\, b_{\mathrm{DAS}}. \tag{15}$$

III. Materials and Methods

A. The Very Large Cardiac Channel Data Database (VLCD)

The Very Large cardiac Channel data Database (VLCD) database was recorded at St. Olavs hospital, Trondheim University Hospital (Trondheim, Norway) in 2018 and 2019. The study was approved by the Regional Committee for Medical & Health Research Ethics in Central Norway. All patients provided written informed consent. All data were recorded by two experienced echocardiographers.

The data were recorded with a GE Vingmed Ultrasound Vivid E95 ultrasound system and the 4Vc-D matrix array probe. The scanner was set up with the native Cardiac_E application and a frequency setting of 1.7/3.4 MHz for second-harmonic imaging. And thus, the reconstructed images in this study are from second harmonic data. The system features ACE (Adaptive Contrast Enhancement) and HD (High Definition) were turned off. The 4Vc-D matrix array probe has an aperture of size 21.5 mm × 15.6 mm mapped to 10 × 19 subaperture (SAP) channels (each channel consists of several pre-beamformed elements), in the azimuth and elevation directions respectively.

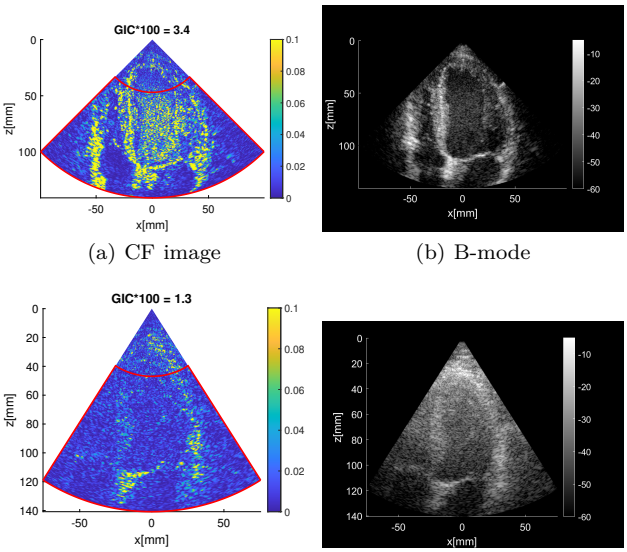
The Vivid E95 has a software beamforming architecture and a special feature on the system, provided to the authors of this article, which allows recording of channel data. These are raw in-phase quadrature (IQ) sampled ultrasound data from each individual SAP prior to general beamforming and image processing. At least one cineloop containing one heart cycle was recorded from the five standard views: parasternal long axis (PLAX), parasternal short axis (PSAX), apical four chamber (A4C), apical two chamber (A2C), and apical long axis (ALAX). Some patients have several recordings of the same cardiac view. For some patients some cardiac views were not recorded due to technical difficulties. A total of 535 channel data recordings were collected from 106 patients containing 33280 individual image frames. On average, there are 64 frames per recording with a frame rate of 40 frames per second (FPS). A typical size of the channel data file is between 2 to 5 GB and the total approximate size of the database is 1.5 TB.

B. Global Image Coherence

A single value of the coherence per image frame, that we denote the Global Image Coherence (GIC) can be calculated from any of the coherence measures described in Section II. Every coherence measure results in one coherence value for every pixel in the image. The GIC is calculated by averaging the pixel coherence values below a pre-defined range R_1 until the end depth of the image R_2. Mathematically, for the CF but valid for all coherence measures, this can be described as

$$\mathrm{GIC}_{\mathrm{CF}} = \frac{1}{\Theta(R_2 - R_1)} \sum_{\theta=1}^{\Theta} \sum_{r=R_1}^{R_2} \mathrm{CF}(\theta, r), \qquad (16)$$

where CF is calculated as in (3), Θ is the number of reconstructed lines in azimuth and $[R_1, R_2]$ is set to $[R_1 = 1/3 \cdot R_{\max}, R_2 = R_{\max}]$ indicated by the red ROI in Fig. 1a and Fig. 1c. Even though an advantage of the GIC is that no manual or segmented ROI is needed, we decided to avoid the region close to the probe to avoid dependency of expanding aperture in the receive apodization when calculating CF, as well as avoiding the top region in the cardiac images which is often quite noisy with potential rib interference and reverberations.



(a) CF image (b) B-mode

(c) CF image (d) B-mode

Fig. 1: The coherence calculated as CF for a frame with high image quality in (a) and lower image quality in (c) with the corresponding b-mode images in (b) and (d) respectively. The red ROI in (a) and (c) indicates the regions used to compute the GIC, selected below a pre-defined range R_1 until the end depth of the image R_2. The GIC is indicated in the title of the image in (a) and (c), notice that the GIC is higher for the image in (a) than in (c).

The coherence calculated as CF is illustrated for a frame with high image quality in Fig. 1a and lower image quality in Fig. 1c with the corresponding b-mode images in Figs. 1b and 1d respectively.

In all the results presented in this paper, the GIC is scaled with 100, as indicated on the applicable axis, for better readability.

C. Conventional Image Quality Contrast Metrics

1) Contrast Ratio (CR): One of the most used measures of contrast in ultrasound imaging is the contrast ratio (CR) [21],

$$\mathrm{CR} = \frac{\mu_{\mathrm{ROI}}}{\mu_{\mathrm{B}}} \qquad (17)$$

where $\mu_{\mathrm{ROI}} = \mathrm{E}\left\{|b_{\mathrm{ROI}}|^2\right\}, \mu_{\mathrm{B}} = \mathrm{E}\left\{|b_{\mathrm{B}}|^2\right\}$, are, respectively, the mean signal power inside a region of interest (ROI) and a background region (B), where b denotes the summed signal from (2). The contrast ratio can take any

positive real value, and CR$\to\infty$ as $\mu_B \to 0$. However, it is often expressed in decibels as,

$$\mathrm{CR[dB]} = 10\log_{10}\mathrm{CR}. \tag{18}$$

2) Generalized Contrast-to-noise-ratio (gCNR): The gCNR was introduced in [15], [22] as a robust and quantitative contrast metric and is calculated using the probability density functions (PDF) of a region of interest (ROI) and a background region (B), for which the contrast, or detectability, is measured, as

$$\mathrm{gCNR} = 1 - \int_{-\infty}^{\infty} \min[p_{\mathrm{ROI}}(x), p_B(x)]dx. \tag{19}$$

This estimates the area of the overlap between the two PDF curves. Two regions with pdfs resulting in a large overlap will have smaller gCNR and detectability compared to two regions with less overlap. This gives a fair and robust comparison between images and was specifically introduced to address the problem of dynamic range alterations [18] in some modern adaptive beamformers but is also valid on conventional DAS images as it is used here.

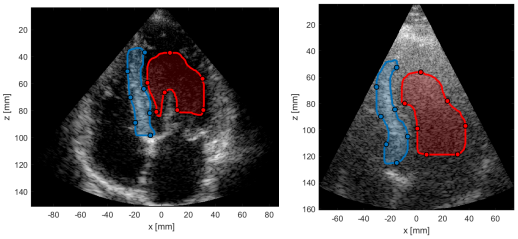
D. Validating GIC using Conventional Contrast Metrics

To validate the GIC we compare and correlate it to the conventional contrast image quality metrics CR and gCNR. However, to estimate the CR and gCNR we need to manually segment an ROI and a background region B. We therefore created a subset of the VLCD containing 20 images of the A4C view. The datasets were selected randomly while making sure we spanned low to high image quality. This was done by selecting the datasets classified as the A4C view, sorting them with GIC ranging from low to high and then uniformly draw 20 datasets from the sorted list of datasets.

The manual segmentation of the ROI and background is done as indicated by high, and one low image quality case in Fig. 2, We chose to segment the major part of the interventricular septum, the heart wall, as the ROI indicated by the blue mask, and compare it with the major part of the left ventricle segmented as the background B indicated by the red mask. The validation is done by correlating the resulting GIC value with the resulting gCNR and CR value from each frame.

E. Data Processing

The beamforming was performed in MATLAB (The Mathworks, Inc., Natick, MA, USA) using the generalized beamformer in the UltraSound ToolBox (USTB, https://www.USTB.no) [23] with retrospective transmit beamforming and a hybrid transmit delay model as described in [24]. To limit the computation time, the beamforming performed for the estimation of the coherence measures used a sector scan of four times the number of transmit directions in the azimuth direction (4 multiple line acquisitions (MLAs)) and 256 depth pixels. For the images displayed in this article, we used 6 MLAs per transmit and 512 depth pixels for a more visually pleasing image.



(a) High quality image (b) Low quality image

Fig. 2: Two of the manually segmented images used to estimate the conventional contrast metrics CR and gCNR to compare agains the results of the GIC. The blue mask is the segmentation of the ROI, the major part of the heart wall, while the red regions, the background B, is the major part of the left ventricle.

F. View classification

In order to classify cardiac views, we used the machine learning based view classifier developed by Østvik et al. [25]. Østvik et al. trained a convolutional neural network using a network structure detailed in [25]. The model was trained and validated using data from 4582 cine loops from 205 patients recorded in transthoracic echocardioagraphy and classified each cine loop into one of seven standard cardiac views A2C (apical two-chamber), A4C (apical four-chamber), ALAX (apical long-axis), PLAX (parasternal long-axis), PSAX (parasternal short-axis), SC4C (subcostal four-chamber), SCVC (subcostal vena cava inferior) as well as a non-assignable class "unknown". The reported accuracy was 98.9±0.6 using cross-validation on the training/validation data. On a separately recorded dataset, a testset, consisting of data from 2559 cine loops of the A4C, A2C, ALAX, PLAX, and PSAX view from 265 patients their model achieved an accuracy of 98.5±0.5. See the discussion for details of the accuracy on our dataset.

G. Statistical Analysis

All statistical analyses were done in MATLAB. To estimate the similarity between coherence measures, and image quality metrics, we calculated the correlation using conventional linear regression and estimation of Pearson's correlation. The statistical difference between the median GIC value for the apical and the parasternal views was tested using a Wilcoxon Rank Sum Test.

IV. Results

The mean value of the GIC averaged over the number of frames in one recording for every coherence measure described in Section II is plotted in the top plot of Fig. 3, with the same but normalized values in the bottom plot. From the plotted values we can visually see a high correlation between the various methods for calculating coherence. The high correlation is confirmed in Fig. 4, where we for every coherence measure calculate

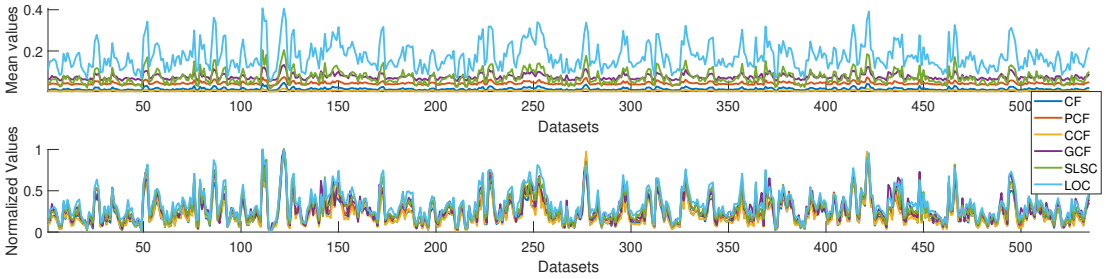


Fig. 3: Mean value over the number of frames for datasets for all coherence measures. The top plot shows raw values from each method, while the bottom shown the values normalized to the same range.

the Pearsons correlation coefficient R, plotted in the lower left triangle of the matrix, while the scatter plot with the estimated regression line is plotted in the upper right triangle.

From the figure we see that all coherence measures correlate with each other with an R value higher than 0.9. With the highest correlation, 0.995 being between the two ways of calculating phase coherence PCF and CCF, and the lowest correlation being between GCF and SLSC. All the p-values are below 0.001 and thus the correlation is statistically significant.

To visually illustrate the amount of data in the VLCD, the mean GIC (using CF) averaged over all frames in each recording is plotted for every dataset in Fig. 5. Two times the standard deviation is indicated by the whiskers in the plot, while the cardiac views, PLAX, PSAX, A4C, A2C and ALAX is indicated in the figure legend. The x-axis shows the patient number as included into the study. Notice that some of the datasets have high GIC (e.g. defined GIC*100>2) , while the greater part of the datasets have rather low GIC (GIC*100<2).

If we sort all the recordings by cardiac view, we can investigate the GIC per view through the box-whisker plot in Fig. 6. The boxplots indicates the median as the red line, the 25th and 75th percentile as the top and bottom of the box, while the notches indicate the 95% confidence interval of the median value and thus we can notice that the apical views have statistically significantly higher GIC values than the parasternal views. The statistical significance is

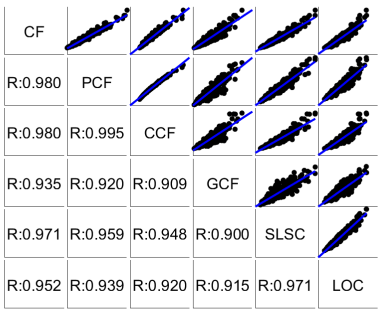


Fig. 4: The upper right of the matrix displays the scatter plots between each coherence measure on all datasets, while the left triangle indicates the resulting Pearson's correlation coefficient. All correlation coefficients have a p-value < 0.001.

confirmed by a Wilcoxon Rank Sum Test with $p < 0.001$.

To obtain a more in-depth analysis of the GIC we have highlighted the GIC per dataset with the cardiac view indicated for patient 21 to patient 24 in Fig. 7a - excluding a double recording of the A2C for Patient 21, and correcting a misclassification of the PLAX as PSAX for patient 22. The corresponding b-mode images displayed with 55 dB dynamic range is shown in Fig. 7b to 7q. Notice that the GIC in (a) corresponds well with the visual interpretation of image quality since the A4C, A2C and ALAX from patient 22 as well as the PSAX, A4C, A2C and ALAX from patient 24 have better image quality than the rest. The same observation is confirmed in Fig. 8 where we plot the GIC and corresponding b-mode images from patients 85 and 86, and we can observe that patient 85 have better image quality than patient 86. A movie loop of the b-mode images are available in the supplementary materials. Since patient 85 did not have a recording of the A2C, we did not display it for patient 86 either.

Fig. 9a plots the GIC against the CR given a segmentation of the ROI and bacground as described in III-D. The plotted line is the estimated linear regression line, which in this case illustrates high correlation between GIC and CR with a Pearson's correlation value of R=0.74 (p<0.005). Notice that the subset of the dataset spans from low to high image quality as measured in terms of GIC. Further, the GIC compared to gCNR is plotted in Fig. 9b, with the line indicating the linear regression results in a fairly high Pearson's correlation value of R=0.62 (p<0.005). Notice that the gCNR saturates with higher image quality, this observation is further elaborated in the discussion.

The GIC for a full recording of low image quality is in Fig. 10, and an image of high image quality is in Fig. 11. The top pane in both plots are the b-mode with the CF images in the middle pane from four frames indicated in the bottom plot. The bottom is the GIC through all 64 frames. The selected frames are to highlight the parts of the cardiac cycle containing the highest GIC. We can observe that for an image of higher quality, the frame reaching the highest GIC is in the diastole phase of the cardiac cycle, where the left ventricle is at its largest due to the high coherence values in the blood. While the lower quality image has less variation in the GIC over the full recording.

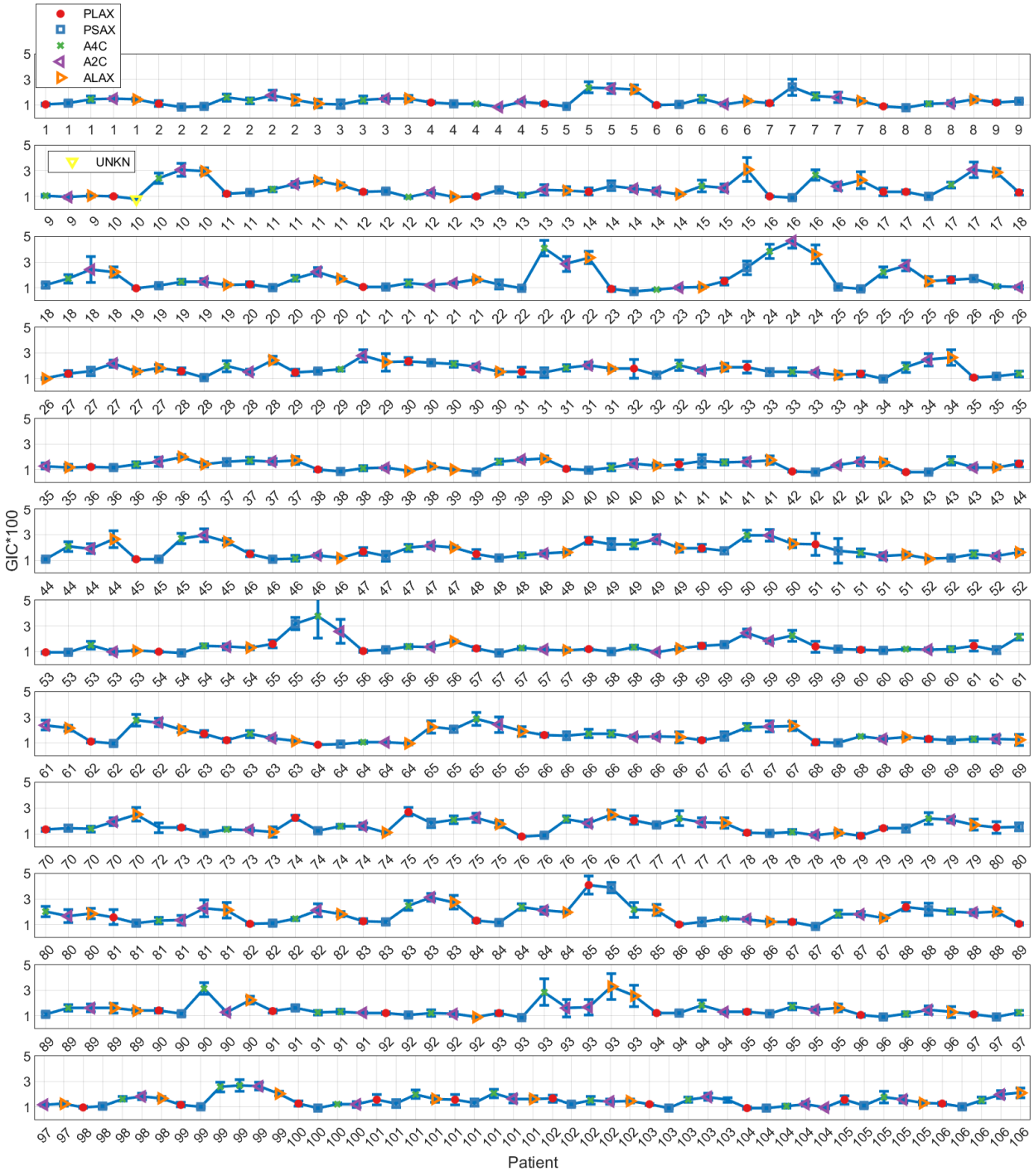


Fig. 5: The mean and two times the standard deviation of the Global Image Coherence (GIC) for all the datasets in the VLCD with view indicated by the marker in the legend. This figure is a graphical illustration of the VLCD.

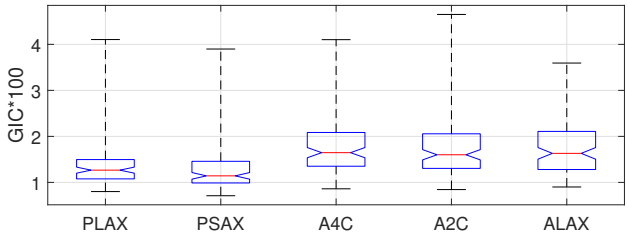
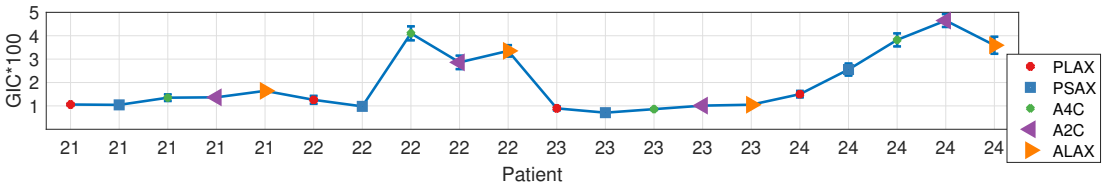
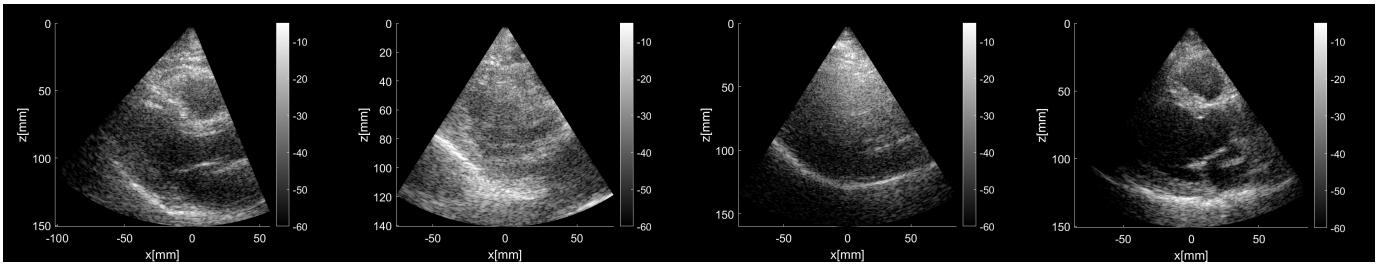


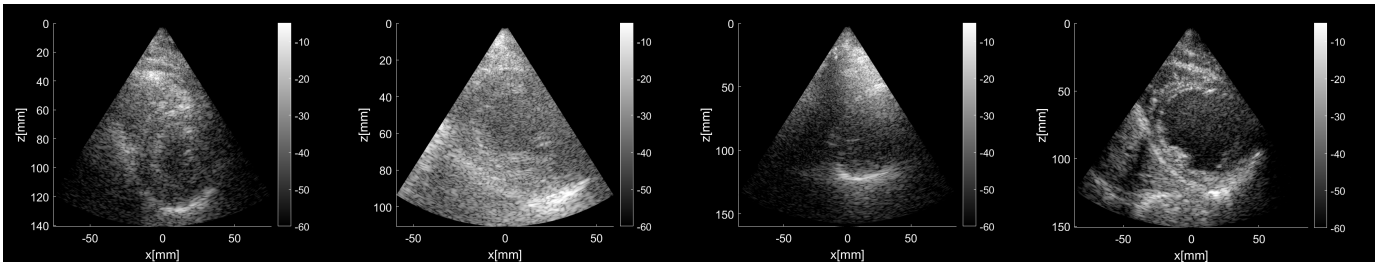
Fig. 6: Boxplot of the GIC from each cardiac view. The red line indicates the median, the bottom and top edge indicated the 25th and 75th percentile while the whiskers are the highest and lowest value. The notches in the box indicate the 95% confidence interval of the median value. Notice that the apical views have statistically significantly higher GIC values than the parasternal views confirmed by a Wilcoxon Rank Sum Test.



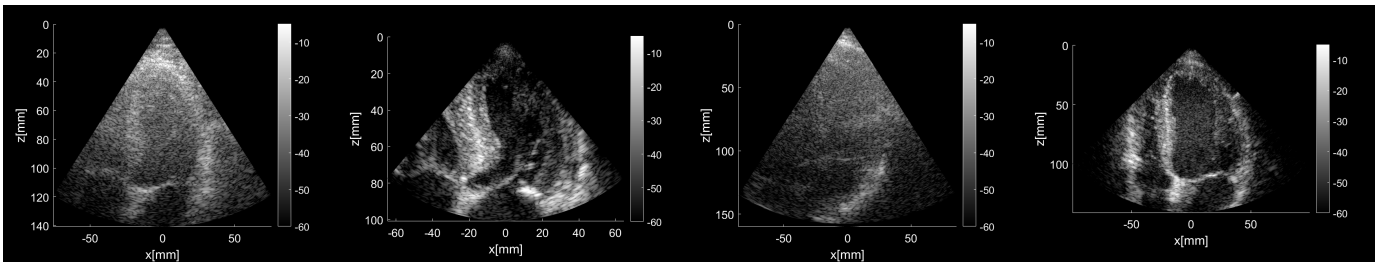
(a) Global Image Coherence (GIC) displayed for patient 21 to 24 with view indicated in the legend. Notice higher GIC for the apical views for patient 22, and all views for patient 24.



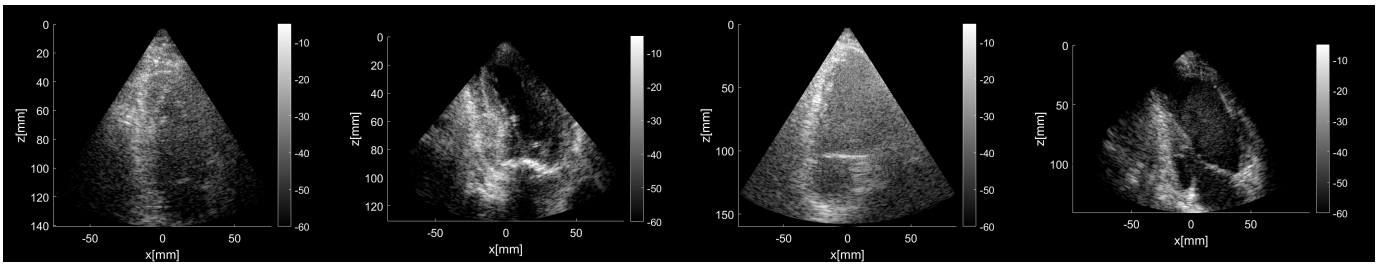
(b) Patient 21 PLAX $\widehat{\text{GIC}} = 1.1$ (c) Patient 22 PLAX $\widehat{\text{GIC}} = 1.3$ (d) Patient 23 PLAX $\widehat{\text{GIC}} = 0.9$ (e) Patient 24 PLAX $\widehat{\text{GIC}} = 1.5$



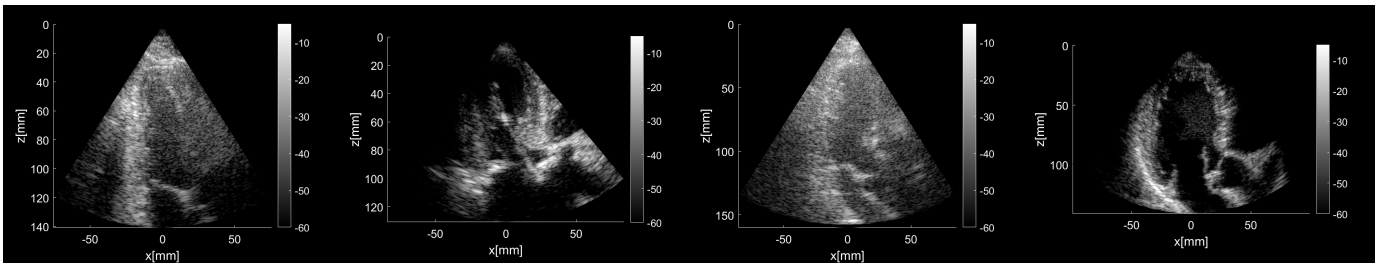
(f) Patient 21 PSAX $\widehat{\text{GIC}} = 1.0$ (g) Patient 22 PSAX $\widehat{\text{GIC}} = 1.0$ (h) Patient 23 PSAX $\widehat{\text{GIC}} = 0.7$ (i) Patient 24 PSAX $\widehat{\text{GIC}} = 2.6$



(j) Patient 21 A4C $\widehat{\text{GIC}} = 1.4$ (k) Patient 22 A4C $\widehat{\text{GIC}} = 4.1$ (l) Patient 23 A4C $\widehat{\text{GIC}} = 0.9$ (m) Patient 24 A4C $\widehat{\text{GIC}} = 3.8$

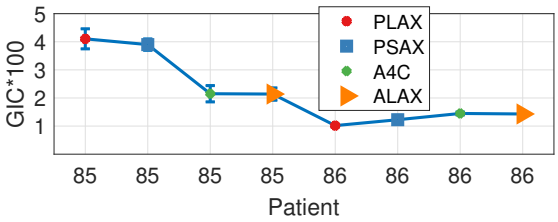


(n) Patient 21 A2C $\widehat{\text{GIC}} = 1.4$ (o) Patient 22 A2C $\widehat{\text{GIC}} = 2.9$ (p) Patient 23 A2C $\widehat{\text{GIC}} = 1.0$ (q) Patient 24 A2C $\widehat{\text{GIC}} = 4.7$

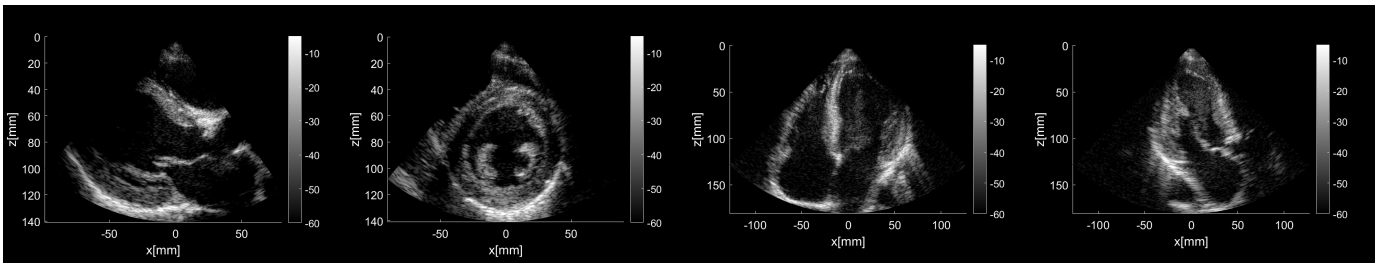


(r) Patient 21 ALAX $\widehat{\text{GIC}} = 1.6$ (s) Patient 22 ALAX $\widehat{\text{GIC}} = 3.4$ (t) Patient 23 ALAX $\widehat{\text{GIC}} = 1.1$ (u) Patient 24 ALAX $\widehat{\text{GIC}} = 3.6$

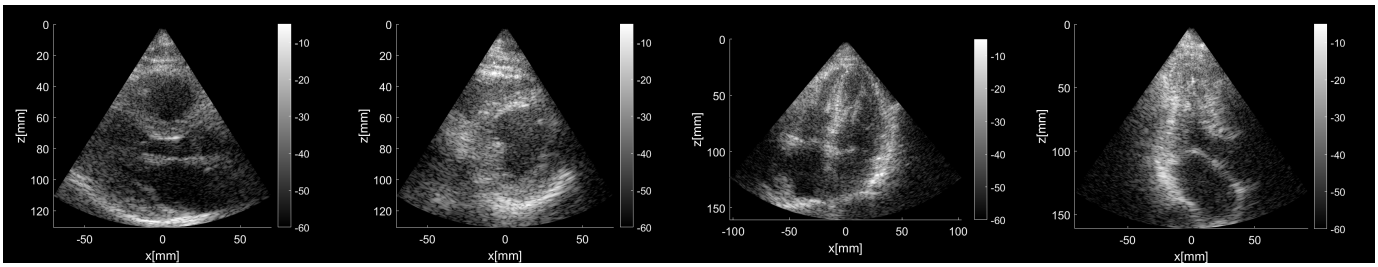
Fig. 7: Fig. (b) to (u) are the b-mode images corresponding to the GIC values plotted in (a). Here $\widehat{\text{GIC}} \equiv \text{GIC} * 100$.



(a) Global Image Coherence (GIC) displayed for patient 85 and 86 with view indicated in the legend. Notice higher GIC for patient 85 compared to patient 86.

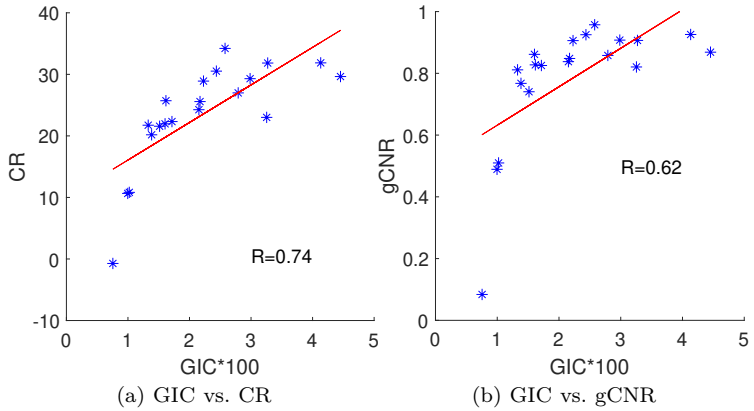


(b) Patient 85 PLAX $\widehat{\text{GIC}} = 4.1$ (c) Patient 85 PSAX $\widehat{\text{GIC}} = 3.9$ (d) Patient 85 A4C $\widehat{\text{GIC}} = 2.2$ (e) Patient 85 ALAX $\widehat{\text{GIC}} = 2.1$



(f) Patient 86 PLAX $\widehat{\text{GIC}} = 1.0$ (g) Patient 86 PSAX $\widehat{\text{GIC}} = 1.2$ (h) Patient 86 A4C $\widehat{\text{GIC}} = 1.4$ (i) Patient 86 ALAX $\widehat{\text{GIC}} = 1.4$

Fig. 8: Fig. (b) to (i) are the b-mode images corresponding to the GIC values plotted in (a). Here $\widehat{\text{GIC}} \equiv \text{GIC} * 100$. A .GIF movie loop of the b-mode images are available in the supplementary materials.



(a) GIC vs. CR (b) GIC vs. gCNR

Fig. 9: In (a) the GIC is plotted against the estimated CR resulting in a correlation value of R=0.74 ($p<0.005$). The GIC compared to gCNR is plotted in (b), results in a correlation value of R=0.62 ($p<0.005$). Notice that the subset of datasets spans from low to high image quality in terms of GIC.

V. Discussion

The results in Fig. 3 and Fig. 4 illustrates that there is a clear correlation between all the inspected coherence measures. Coherence is impacted by local effects including backscatter contrast, absolute scattering strength, clutter level, focal effects, local aberration profile. These effects might lead to local variations within a single frame between the different coherence measures we have investigated. However, when we average these quantities over an entire image frame and then over all frames in a recording, as we do in our evaluation of the GIC, our results indicates that although there are differences in how the coherence measures are implemented and their absolute values, the information content is very similar. We can therefore confirm our first hypothesis that published coherence measures are strongly correlated.

The second hypothesis, that GIC can be used as a quantitative metric for in-vivo image quality can be visually confirmed for cardiac images when we compare the measured value of the GIC to the b-mode images in Fig. 7a and 8. This is in agreement with the findings in [8] where they derived an analytic expression relating coherence through LOC to the channel signal-to-noise ratio. However, when we sort the GIC values by view in Fig. 6 we see that there is a significant difference in the GIC value between the parasternal and the apical views, with the apical views reaching a higher GIC value. This could be because the amount of tissue visible in the image is different in the apical and parasternal views. In further works one should aim at normalizing the GIC on the amount of tissue in the images to obtain a GIC value that can be quantitatively compared across any cardiac view. The GIC difference between the apical and parasternal views could also be that the parasternal views are more prone to the ribs blocking the aperture or the lung interfering with the image quality. However, further analysis is needed to explain the difference in GIC between the apical and parasternal views.

When we investigate the GIC per frame in a full recording for a low quality recording in Fig. 10 and a high quality recording in Fig. 11 we noticed that frames in the diastole phase had the highest GIC for the recording with high image quality. This is probably because in the high quality recordings the blood speckles are visible in the b-mode image. Since blood is coherent it is, together with tissue, contributing to the higher GIC. While in the lower quality recording, the blood speckles are not visible in the b-mode image and thus does not contribute to the GIC value. We can also observe in Fig. 5 that higher quality images (higher GIC) seems to have a higher variance of GIC over frames in a recording. This is most likely since high quality recordings have blood contributing to the GIC value, and the amount of blood in the image varies through the cardiac cycle.

The classification accuracy of the view classificator on the test set in [25] was $98.5 \pm 0.5\%$. Our dataset consists of the same cardiac views as in their test set, also recorded on clinical systems from the same vendor GE Vingmed Ultrasound. We therefore assume our dataset to be classified with a similar accuracy even though we have not manually gone through all our 538 datasets. As shown in 5 there was only one dataset resulting in the unknown category (one dataset for patient 10). We can also notice that this dataset was of low quality as indicated by the low GIC value. From 5 we can also verify that almost all patients had all 5 cardiac views as according to the protocol. Some patients, for example patient two, have multiple recordings of some of the views. While some patients do not have all views, for example patient four lack the ALAX view. The missing views are most likely due to technical difficulties while storing the raw channel data. One misclassification was observed for patient 22 where the PLAX was incorrectly classified as PSAX. This was corrected from Fig. 5 to Fig. 7a. However, we can once again notice from the GIC value and the images in Fig.7c and 7g that the misclassification is most likely due to the low image quality.

Further confirming our second hypothesis, the validation of the GIC against the conventional contrast metrics CR and gCNR resulting in the plots in Fig. 9 illustrate that we have a high correlation against the conventional contrast metrics. However, there is a higher correlation with CR ($R=0.74$, $p<0.005$) than with gCNR ($R=0.62$,$p<0.005$). An explanation for this could be that from Fig. 9b, we can notice that the gCNR saturates for images with higher image quality. This is probably explained by the fact that gCNR estimates image quality in terms of detectability. The ROI, in our case the interventricular septum, is easily detected from the background, the left ventricle, when the image quality reaches a certain level - and thus the ROI and the bacground well separated also at medium image quality. Having an image of higher image quality does not necessarily increase the separability. Based on this it can be debated how well the gCNR is suited as an quantitative image quality metric for different in-vivo cardiac images, and thus how relevant separability between two ROIs are as an in-vivo metric. The gCNR was first and foremost introduced to evaluate advanced and non-linear beamforming methods comparing beamforming methods applied to the same data. Here we are using it to compare quality across different datasets. This adds to the discussion that quantitatively evaluating image quality in in-vivo ultrasound images is hard, and that we need a collection of metrics to do fair and correct evaluations. We believe that the GIC is a valuable contribution to these image quality metrics.

A more relevant validation of the GIC would be a clinical validation with e.g. clinicians ranking images in terms of image quality and estimating how well this corresponds to the GIC value. Then one could also assess if there are quantitative bounds on image quality, for example if the

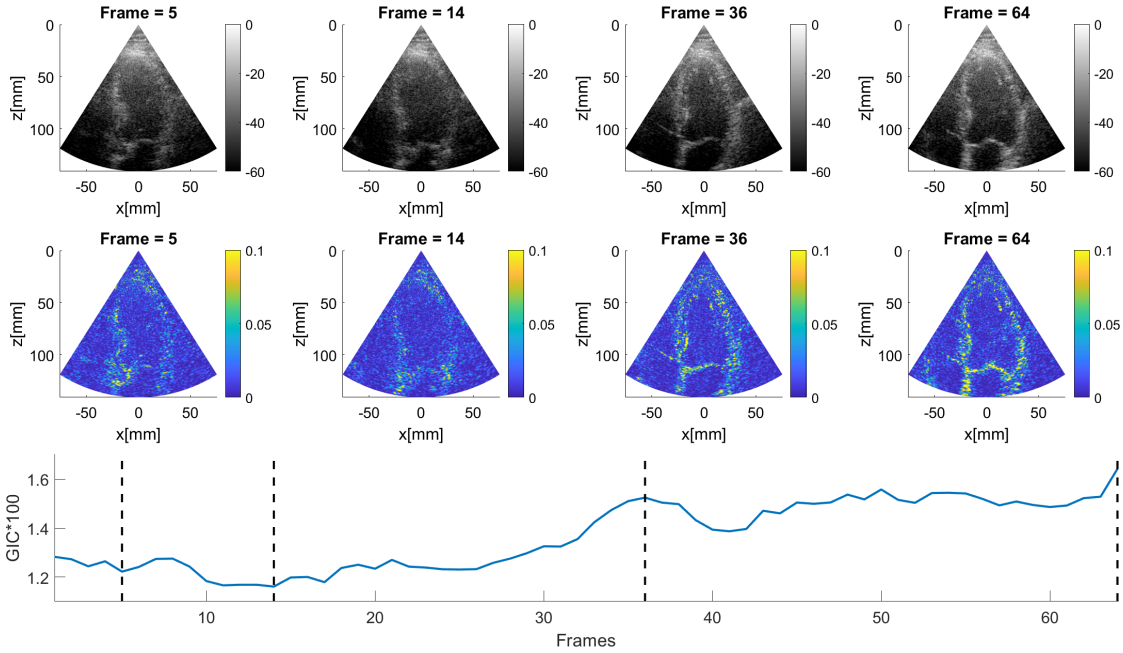


Fig. 10: The top pane are the b-mode with the coherence images in the middle pane from four frames indicated in the bottom plot. The bottom is the GIC through all 64 frames of a lower quality image. The selected frames illustrate frames with lower and higher GIC.

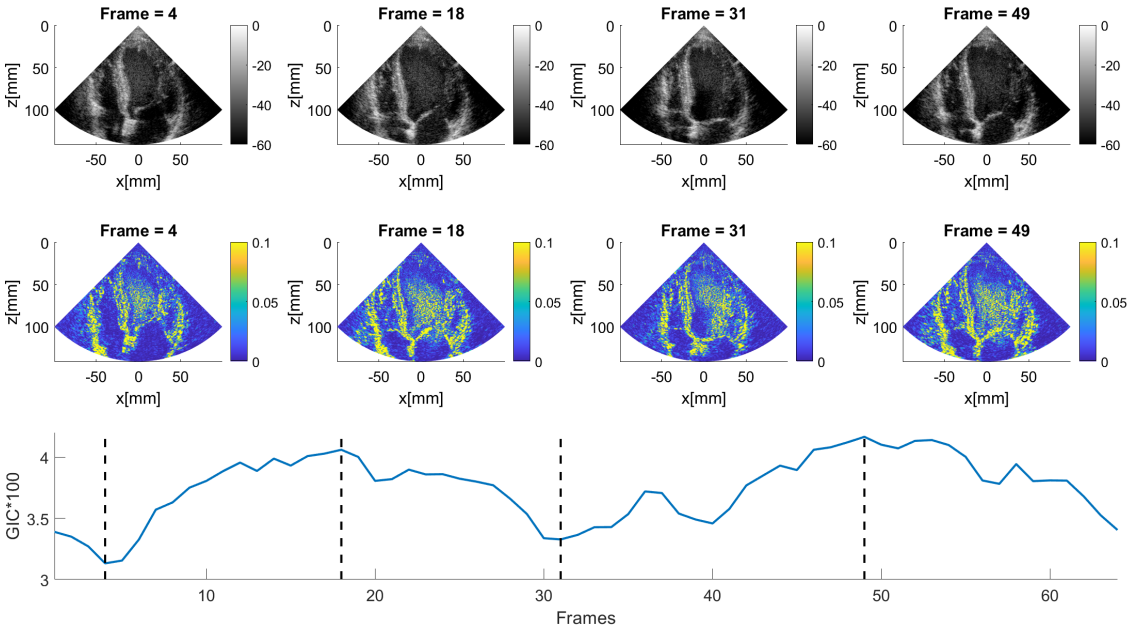


Fig. 11: The top pane are the b-mode with the coherence images in the middle pane from four frames indicated in the bottom plot. The bottom is the GIC through all 64 frames of a higher quality image. The selected frames illustrate frames with lower and higher GIC.

GIC quantitatively can support classifications into low-, medium- and high image quality. Such a study should be performed. However, we have recently published a relevant study in [17] where we used a variant of the GIC to estimate the image quality improvements resulting from an aberration correction algorithm. This was a clinical study where four clinicians evaluated cardiac cineloops with and without aberration correction blinded, left-right-randomized and side-by-side. The clinicians' task was to choose which image they preferred and the aberration corrected image was preferred in 97% of the cases of the 116 recorded cardiac cineloops. This matched well with

the values of the GIC which showed an increase in GIC in all frames after aberration correction.

In this work, we have focused on cardiac ultrasound images which may be advantageous to the GIC since they have well defined standard scan views. This allows for easier comparison of GIC between recordings since the clinicians have strived to obtain the same view in all recordings. In other clinical ultrasound applications that does not have as standardized views the GIC will still be valuable to e.g. optimize beamforming parameters for specific recordings, but it might be harder to compare recordings from different patients.

A possible clinical utility of these results is a case where a clinician is recording several cineloops of a patient, and one of the cineloops should be selected for further post processing and estimation of clinical cardiac diagnostic parameteres. Automatic selection of the cineloop with highest GIC value may improve diagnostic accuracy, and we hypothesise that automatically selecting the cine loop with the highest image quality will further support automatic guidance with deep learning tools [26], [27]. The GIC could also indicate the reliability of measurements made in a certain image. This is useful both while recording and doing the measurements, but perhaps even more important reviewing a list of measurement made on a recording. Also, having a quantitative measure of in-vivo image quality can allow automatic tuning of parameters such as transmit setups [8] and other beamforming parameters to achieve an image of the highest image quality for every patient scanned. Perhaps this can tilt the odds in favour of the clinician in the lottery of cardiac image quality. Therefore, the GIC can strengthen echocardiographic diagnostics, and ultimately improve the care of cardiac patients.

VI. Conclusion

The acquisition of the Very Large Cardiac channel data Database (VLCD) consisting of 33 280 individual frames from 538 recordings of 106 patients allowed us to do an empirical study of coherence as an in-vivo image quality metric. We demonstrate that all the coherence measures investigated in this study are highly correlated ($R>0.9$, $p<0.001$) across the database, illustrating that even though there are differences in the implementation and absolute values of the coherence measures, their information content is very similar. Any coherence measure can be averaged across the pixels in an image frame into a Global Image Coherence (GIC) and we demonstrated empirically that this value can be used as a quantitative value for in-vivo image quality. We validated the GIC against the conventional contrast metrics contrast ratio (CR) and the generalized contrast-to-noise ratio (gCNR) on a subset of the full dataset, and obtained a correlation of $R=0.74$ ($p<0.005$) and $R=0.62$ ($p<0.005$) respectively demonstrating a high correlation against conventional contrast metrics. We used the coherence factor (CF) to implement the GIC, however, the exact choice of coherence measure is probably not critical.

Acknowledgment

We would like to thank David Pasdeloup for developing code allowing us to easily use the view classificator developed by Østvik et. al. [25]. We would also like to thank the anonymous reviewers for constructive comments improving the manuscript.

References

[1] R. M. Lang, L. P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F. A. Flachskampf, E. Foster, S. A. Goldstein, T. Kuznetsova, P. Lancellotti, D. Muraru, M. H. Picard, E. R. Rietzschel, L. Rudski, K. T. Spencer, W. Tsang, and J. U. Voigt, "Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American society of echocardiography and the European association of cardiovascular imaging," European Heart Journal Cardiovascular Imaging, vol. 16, no. 3, pp. 233–271, 2015.

[2] P. C. Li and M. L. Li, "Adaptive imaging using the generalized coherence factor," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 50, no. 2, pp. 128–141, 2003.

[3] J. Camacho, M. Parrilla, and C. Fritsch, "Phase coherence imaging," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 56, no. 5, pp. 958–974, 2009.

[4] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Short-lag spatial coherence of backscattered echoes: Imaging characteristics," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 58, no. 7, pp. 1377–1388, 2011.

[5] J.-F. Synnevåg, A. Austeng, and S. Holm, "Benefits of minimum-variance beamforming in medical ultrasound imaging," Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on, vol. 56, no. 9, pp. 1868–1879, 2009.

[6] G. Matrone, A. S. Savoia, G. Caliano, S. Member, and G. Magenes, "The Delay Multiply and Sum Beamforming Algorithm in Ultrasound B - Mode Medical Imaging," IEEE Trans. Med. Imaging, vol. 34, no. 4, pp. 1–10, 2015.

[7] B. M. Asl and A. Mahloojifar, "Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 57, no. 11, pp. 2381–2390, 2010.

[8] W. Long, N. Bottenus, and G. E. Trahey, "Lag-One Coherence as a Metric for Ultrasonic Image Quality," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 65, no. 10, pp. 1768–1780, 2018.

[9] R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert Zernike approach and focusing criterion," The Journal of the Acoustical Society of America, vol. 96, no. 6, p. 3721, 1994.

[10] W. F. Walker and G. E. Trahey, "Speckle coherence and implications for adaptive imaging," The Journal of the Acoustical Society of America, vol. 101, no. 4, pp. 1847–1858, 1997.

[11] N. B. Bottenus and G. E. Trahey, "Equivalence of time and aperture domain additive noise in ultrasound coherence," The Journal of the Acoustical Society of America, vol. 137, no. 1, pp. 132–138, 2015. [Online]. Available: http://dx.doi.org/10.1121/1.4904530

[12] G. Pinton, G. Trahey, and J. Dahl, "Spatial coherence in human tissue: Implications for imaging and measurement," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 61, no. 12, pp. 1976–1987, 2014.

[13] S. W. Smith, H. Lopez, and W. J. Bodine, "Frequency independent ultrasound contrast-detail analysis," Ultrasound in Medicine and Biology, vol. 11, no. 3, pp. 467–477, 1985.

[14] M. S. Patterson and F. S. Foster, "Improvement and Quantitative Assessment of B-Mode Images Produced By Annular Array/Cone Hybrids." Acoustical Imaging: Proceedings of the International Symposium, vol. 13, p. 477, 1984.

[15] A. Rodriguez-Molares, O. M. H. Rindal, J. D'Hooge, S. E. Masoy, A. Austeng, M. A. Lediju Bell, and H. Torp, "The Generalized Contrast-to-Noise Ratio: A Formal Definition for Lesion Detectability," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 4, pp. 745–759, 2020.

[16] O. M. H. Rindal, A. Austeng, A. Fatemi, and A. Rodriguez-Molares, "The Effect of Dynamic Range Alterations in the Estimation of Contrast," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 66, no. 7, pp. 1198–1208, 2019.
[17] S.-E. Masoy, B. Denarie, A. Sornes, E. Holte, B. Grenne, T. Espeland, E. A. R. Berg, O. M. H. Rindal, W. Rigby, and T. Bjastad, "Aberration correction in 2D echocardiography," Quantitative Imaging in Medicine and Surgery, pp. 1–11, 2023. [Online]. Available: https://dx.doi.org/10.21037/qims-22-895
[18] J. Camacho and C. Fritsch, "Phase coherence imaging of grained materials," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 58, no. 5, pp. 1006–1015, 2011.
[19] D. Hyun, G. Trahey, M. Jakovljevic, and J. Dahl, "Short-lag spatial coherence imaging on matrix arrays, Part 1: Beamforming methods and simulation studies," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 61, no. 7, pp. 1101–1112, 2014.
[20] R. Mallart and M. Fink, "The van Cittert-Zernike theorem in pulse echo measurements," Journal of Acustical Society of America, vol. 90, no. November 1991, pp. 2718–2727, 1991.
[21] S. Smith, H. Lopez, and W. Bodine, "Frequency independent ultrasound contrast-detail analysis," Ultrasound in Medicine and Biology, vol. 11, no. 3, pp. 467 – 477, 1985. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0301562985901589
[22] A. Rodriguez-Molares, O. M. H. Rindal, J. D'hooge, S.-E. Måsøy, A. Austeng, and H. Torp, "The Generalized Contrast-to-Noise Ratio," IEEE International Ultrasonics Symposium, IUS, no. 6, pp. 1–4, 2018.
[23] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, A. Nair, M. A. Lediju Bell, H. Liebgott, A. Austeng, and L. Løvstakken, "The UltraSound ToolBox," IEEE International Ultrasonics Symposium, IUS, pp. 1–4, 2017.
[24] O. M. H. Rindal, A. Rodriguez-Molares, and A. Austeng, "A simple, artifact-free, virtual source model," IEEE International Ultrasonics Symposium, IUS, pp. 1–4, 2018.
[25] A. Østvik, E. Smistad, S. A. Aase, B. O. Haugen, and L. Løvstakken, "Real-Time Standard View Classification in Transthoracic Echocardiography Using Convolution Neural Networks," Ultrasound in medicine & biology, vol. 45, no. 2, pp. 374–384, 2019.
[26] D. Pasdeloup, S. H. Olaisen, A. Østvik, S. Sabo, H. N. Pettersen, E. Holte, B. Grenne, S. B. Stølen, E. Smistad, S. A. Aase, H. Dalen, and L. Løvstakken, "Real-Time Echocardiography Guidance for Optimized Apical Standard Views," Ultrasound in Medicine and Biology, vol. 49, no. 1, pp. 333–346, 2023.
[27] S. Saeboe, H. N. Pettersen, A. Oestvik, D. Pasdeloup, E. Smistad, S. Stoelen, B. Grenne, L. Loevstakken, E. Holte, and H. Dalen, "Automated analyses and real-time guiding by deep learning to reduce test-retest variability of global longitudinal strain," European Heart Journal, vol. 43, 10 2022.

Ole Marius Hoel Rindal was born in Hamar, Norway in 1990. He received the M.S degree in computer science (signal processing) in 2014 from the University of Oslo (UiO), Norway, and in 2019 he received the Ph.D. degree with a dissertation entitled "Software Beamforming in Medical Ultrasound Imaging - a blessing and a curse". He is currently a postodoc and senior lecturer at UiO as well as a co-founder and technology manager in the company Sonair AS. His research interest includes signal processing, beamforming, medical image formation, image processing, machine learning, as well as physiological and biomechanical aspects of cross-country skiing.

Tore Bjåstad (M'05) was born in Aalesund, Norway, in 1977. He received the sivil ingenioer (M.Sc.) degree in electronics and telecommunications from the Norwegian University of Science and Technology (NTNU) in 2002 and his Ph.D. degree from the Department of Circulation and Medical Imaging, Faculty of Medicine, in 2009. He is currently employed at GE Healthcare, GE Vingmed Ultrasound AS. His main research interests are beamforming and signal processing methods for improved ultrasound image quality.

Torvald Espeland was born in Stord, Norway, in 1981. He received his Medical Degree from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2008. He is currently working as a Ph.D. candidate at the Department of Circulation and Medical Imaging at NTNU and as a Medical Doctor at the Clinic of Cardiology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway. His main research interests are cardiac imaging, advanced echocardiography and cardiovascular physiology and hemodynamics.

Erik Andreas Rye Berg was born in Bergen in 1981. He graduated with his Medical Degree from the Norwegian University of Science and Technology (NTNU) in 2008. He has performed research at Centre for Innovative Ultrasound Solutions and NTNU since 2016. Since 2017 he has worked as consultant in internal medicine/cardiology at St. Olavs hospital, Orkdal, Norway. He currently holds a position as fellow in cardiology at St. Olavs hospital, Trondheim, Norway.

Svein-Erik Masøy was born in Halden, Norway, in 1975. He received his M.Sc. degree in applied mechanics, thermodynamics, and fluid dynamics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 1999. In 2004, he received the Ph.D. degree with his dissertation entitled Estimation and correction of aberration in medical ultrasound imaging. He is an assosiate professor at the Department of Circulation and Medical Imaging at NTNU as well as center director for the Center for Innovative Ultrasound Solutions (CIUS). His research interests are nonlinear acoustics, beamforming, machine learning, and non-destructive testing using ultrasound.