Automated Aortic Doppler Flow Tracing for Reproducible Research and Clinical Measurements

Massoud Zolgharni*, Niti M. Dhutia, Graham D. Cole, M. Reza Bahmanyar, *Member, IEEE*, Siana Jones, S. M. Afzal Sohaib, Sarah B. Tai, Keith Willson, Judith A. Finegold, and Darrel P. Francis

***Abstract*—In clinical practice, echocardiographers are often unkeen to make the significant time investment to make additional multiple measurements of Doppler velocity. Main hurdle to obtaining multiple measurements is the time required to manually trace a series of Doppler traces. To make it easier to analyze more beats, we present the description of an application system for automated aortic Doppler envelope quantification, compatible with a range of hardware platforms. It analyses long Doppler strips, spanning many heartbeats, and does not require electrocardiogram to separate individual beats. We tested its measurement of velocity-time-integral and peak-velocity against the reference standard defined as the average of three experts who each made three separate measurements. The automated measurements of velocity-time-integral showed strong correspondence ($R^2 = 0.94$) and good Bland–Altman agreement ($SD = 1.39$ cm) with the reference consensus expert values, and indeed performed as well as the individual experts ($R^2 = 0.90$ to 0.96, $SD = 1.05$ to 1.53 cm). The same performance was observed for peak-velocities; ($R^2 = 0.98$, $SD = 3.07$ cm/s) and ($R^2 = 0.93$ to 0.98, $SD = 2.96$ to 5.18 cm/s). This automated technology allows > 10 times as many beats to be analyzed compared to the conventional manual approach. This would make clinical and research protocols more precise for the same operator effort.**

***Index Terms*—Doppler measurements, echocardiography, ultrasound imaging.**

I. Introduction

DOPPLER echocardiography is the gold-standard clinical method for measurement of blood velocity in the heart and great vessels [1], and is a cornerstone in the assessment of valvular heart disease and cardiac performance. It requires a skilled operator to acquire the images and, subsequently, analyze them by manually tracing around the Doppler envelopes to measure velocity-time-integral (VTI) and peak velocity. The shape and size of the Doppler envelope varies from beat to beat and is also affected by probe and sample volume position. Echocardiographers considering making multiple measurements to obtain an average have to justify this extra time expenditure or disinvest attention in other areas of the study. When averaging is needed, the echocardiographers tend to select a representative beat which they consider an average beat. We have shown that this may contribute to the significant test–retest variability of Doppler measurements [2].

The ability to acquire and analyze large number of beats would permit clinical protocols to be developed to reduce undesirable variability between clinical assessments, but such an ability would depend on automatic quantification because the labour of tracing large number of heartbeats would be prohibitive. There are several challenges to proving automatic quantification, even for the simplest trace which is the left ventricular outflow tract (LVOT), a location to measure the flow into aorta.

A. Need to Remove all Manual Intervention

Partial automation is beneficial, but if some manual input is still required per beat or per script, then this limits the ability of the operator to focus their time on acquiring large amounts of high quality data. Often forgotten is the extra step of transferring the recordings of ultrasound machine into a separate computer for processing, which is simple but requires some effort and prevents the analysis from being almost real-time.

There have been studies addressing the automatic or semi-automatic tracing of Doppler envelopes, mainly based on the noise-removal and edge-detection techniques [3]–[8]. Learning-based and probabilistic-framework algorithms for automatic detection and segmentation of the deformable Doppler have also been reported [9], [10]. In a recent approach, contour detection method was proposed and applied to Doppler echocardiographic velocity measurements [11]. Most of these reports have not focused on eliminating all manual steps, and require other collateral information such as electrocardiogram (ECG).

B. Potential Value of Longer Recordings

Existing approaches have not been targeted for applications where long uninterrupted recordings are specifically useful. For instance, the fine tuning of cardiac pacemakers where detecting a subtle change in the mean value amongst much larger background beat-to-beat variability is essential for reliable selection of the correct pacemaker setting.

Manuscript received November 18, 2013; revised January 23, 2014; accepted January 26, 2014. Date of publication January 30, 2014; date of current version April 22, 2014. This work was supported by the European Research Council (281524). *Asterisk indicates corresponding author.*

*M. Zolgharni is with the National Heart and Lung Institute, Imperial College London, W12 0NN London, U.K. (e-mail: massoud@zolgharni.com).

N. M. Dhutia, G. D. Cole, S. Jones, S. M. A. Sohaib, S. B. Tai, K. Willson, J. A. Finegold, and D. P. Francis are with the National Heart and Lung Institute, Imperial College London, W12 0NN London, U.K. (e-mail: niti.dhutia04@imperial.ac.uk; g.cole@imperial.ac.uk; siana.jones@imperial.ac.uk; sarah.tai09@imperial.ac.uk; k.willson98@imperial.ac.uk; d.francis@imperial.ac.uk; afzalsohaib@hotmail.com; judyfinegold@doctors.org.uk).

M. R. Bahmanyar is with the Centre for Bio-Inspired Technology, Institute of Biomedical Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: m.bahmanyar@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2014.2303782

Recordings of order of 30 s have shown useful physiological information albeit with laborious manual tracing [12], and even longer recordings would allow probing of the subsequent physiological reactions to intervention. We have shown that for determining the optimal setting of a resynchronization pacemaker, the number of heartbeats that would need to be averaged is well over a hundred [13]–[15].

In research scenarios, it can be important to be able to state in a reliable manner whether stroke volume has changed from one visit to another, using noninvasive techniques. Making measurements with nearly a handful of heartbeats results in test–retest variability that makes it impossible to detect improvement unless a very large number of patients are studied.

In clinical practice, the standard approach for quantifying severity in aortic valve stenosis relies on Doppler assessment of aortic valve and LVOT. If the test–retest variability between visits is large, it is not possible to reliably tell whether patient has been deteriorated or not [2]. It has been shown that the assessment variability of aortic stenosis severities by Doppler echocardiography is high (28%–41%) [16], and the method needs corrective actions [17].

This variability could potentially be reduced by making multiple acquisitions with the probe repeatedly repositioned to obtain a subtly different estimate each time. The average of these estimates would be much more consistent between visits than a single beat in a single position. Such a protocol is impractical when there is any manual work necessary for envelope analysis.

C. Potential Value of Independence From ECG

Cardiac timing is usually provided by obtaining an ECG during image acquisition. However, it may not be convenient to connect ECG cables, particularly in an era when highly portable scanners may be used to undertake focused studies lasting only a few minutes [18].

There have been a few recent studies on ECG-free cardiac cycle detection. In the absence of ECG signal, tissue Doppler data has been used to calculate a gating signal that can be used for dynamic 3-D reconstructions of the foetal heart [19]. Automatic detection of end-diastole and end-systole frames has also been reported by applying manifold learning techniques to 2-D echocardiography images, and for the calculation of the ejection fraction [20]. In another study, apical B-mode (2-D) recordings were used for the automatic detection of cardiac cycle length and cardiac cycle starting time-point [21].

We investigated the feasibility of estimating the cardiac cycle length from Doppler traces, without using the ECG data. Although having the ECG data provides the possibility of computing some parameters of clinical importance such as temporal intervals from the R-wave peaks, we believe that the capacity of estimating the cardiac cycle length independent from the ECG signal could potentially be very useful for implementing the automated technology on the handheld devices in which obtaining the ECG data is not convenient. However, our proposed methods do not prevent collecting the ECG data, if needed.

D. Aims

In this paper, we aim to present the description of an application system to address the clinical need for automated aortic Doppler measurements. LVOT Doppler traces of any desired length are captured from the video output of the echo hardware, and peak velocity and VTI are measured for each beat. This allows the measurement results to be available to the operator within a few seconds.

II. Methods

A. Patient Population

Pulsed wave Doppler images were collected from 18 patients (nine males), with mean age of 59.3, who were referred for echocardiographic examination in the Echocardiography Department at St Mary's Hospital in November 2012 (15 patients in sinus rhythm and three in atrial fibrillation). There were no selection criteria, and none of the patients had aortic stenosis. The study was approved by the local ethics committee and written-informed consent was obtained from all patients.

B. Data Collection

Transthoracic Doppler echocardiography was performed using a GE Vivid.i (GE Healthcare, U.K.) ultrasound machine equipped with a 1.5–3.6 MHz transducer (3S-RS). The sample size was set to 4.9 mm. For each series of Doppler traces, an apical five chamber view was obtained first. The pulsed-wave Doppler cursor was then positioned in the LVOT as recommended in guidelines [22]. The operators performing the exam were instructed not to change the machine settings (e.g., gain, axis scaling, baseline, etc.) during the acquisition period (30 s) in order to obtain a continuous stream of Doppler traces.

Frames, with the original color depth and resolution displayed on the screen (here, 800×600), were captured using a video frame grabber (VGA2USB Pro, Epiphan Systems, Canada), connected to the VGA output on the ultrasound machine, and saved onto a laptop via the USB port [Fig. 1(a)]. The frame grabber reads the data from the analog VGA signal and converts it into a digital RGB image which is provided as a 1-D vector of pixel values. The frame grabber can also be connected to the ultrasound machines with other types of video interface (e.g., DVI, HDMI). The frames were collected at a minimum frame rate of 40 frames per second which was higher than the refresh rate on the screen and the VGA output (30 Hz), making sure no frame was dropped.

In order to test the concept that the hardware and software were independent of particular vendors, Doppler data was also collected from two other widely-used ultrasound machines which were accessible to us, one patient for each, available at St. Mary's Hospital; PHILIPS iE33 xMATRIX (Philips Healthcare, U.K.) and Esaote MyLabTMTwice (Esaote, U.K.). The results are provided in Appendix A. We found that no changes were required for the hardware or software.

C. Feature Extraction From Doppler Snapshots

A rectangular segment of each image frame (hereinafter referred to as "Doppler-region") is used to display the Doppler traces [Fig. 1(b)]. The Doppler-region, outside of which other information not relevant to Doppler traces appear, must first be isolated from the surrounding features. Each echo hardware has alternative positions and sizes for the Doppler-region that the

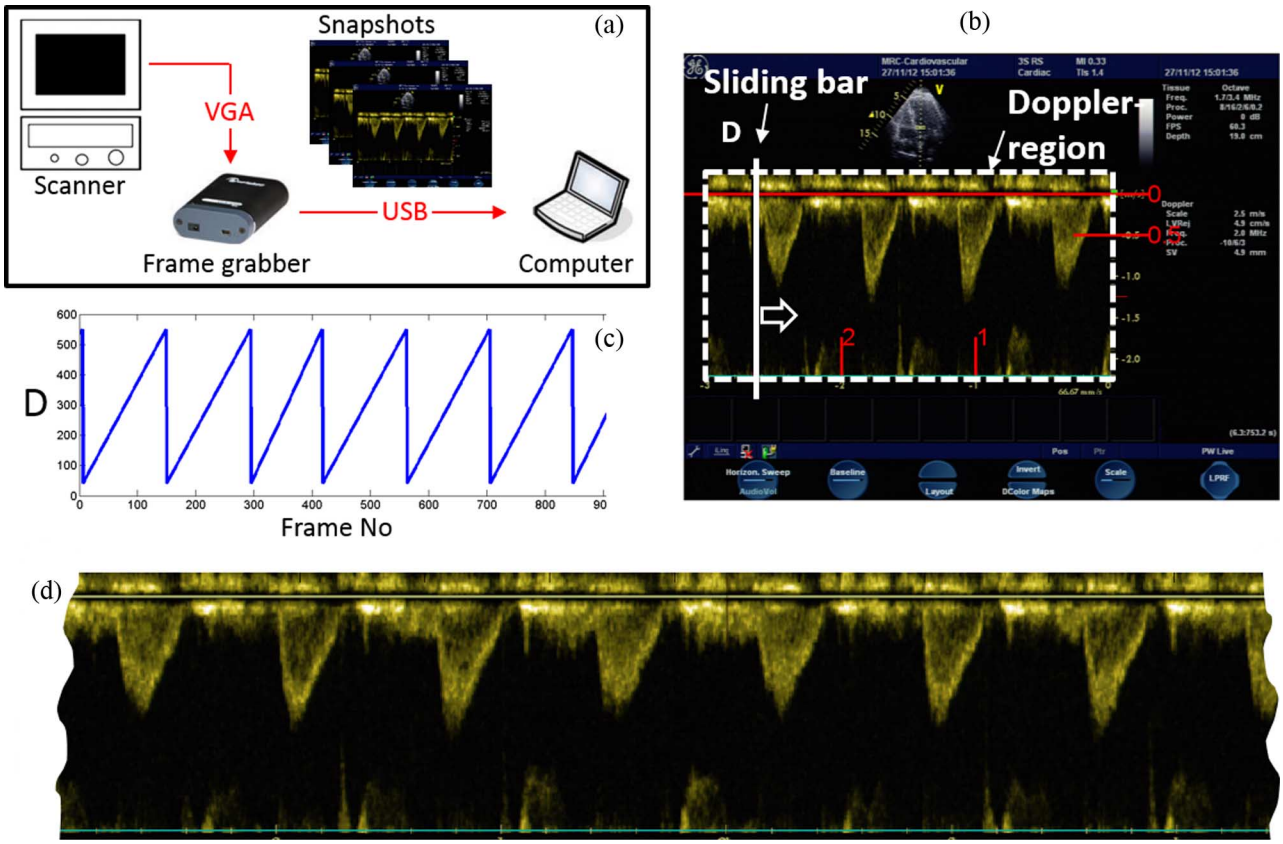

Fig. 1. Different stages of data collection: (a) schematic setup for collection of snapshots of still-frames, (b) automated extraction of Doppler-region, zero-velocity axis, scaling ratios, and updating sliding bar, (c) an example plot of automatically detected location of updating sliding bar where D in pixels shows the column number on the captured image, (d) a cut of a long Doppler strip created by splicing sequential frames.

operator can select from; there are four predefined layouts on GE Vivid.i machines. We built a bank of templates for these layouts that can be used, together with template matching techniques, to automatically crop the captured frames and extract the Doppler-region for each dataset.

Similarly, the scaling ratios, for converting pixel to velocity on the vertical axis and pixel to time on the horizontal axis, can be calculated by extracting templates of different numbers for each echo machine with its unique font style, size, and color. Once the boundaries of the Doppler-region is determined, the numbers are sought for along the boundaries using the template matching techniques. We found this approach more reliable than using the general-purpose optical character recognition techniques. Fig. 1(b) shows the location of automatically detected unit marks and numbers on velocity and time axes. The position of zero-velocity horizontal line can also be identified by averaging the pixel values along each row in the Doppler-region and selecting the maximum. The zero-velocity line was later used to separate the negative LVOT velocities for further processing.

On most echo machines, the Doppler waveform appears in the Doppler-region where the velocity data is updated between consecutive frames by sweeping the Doppler image via a sliding bar from left to right [Fig. 1(b)]. Each frame contains a mixture of new information and information contained in earlier frames. The location of this sliding bar on each frame is detected by comparing two consecutive frames. Apart from the area in the vicinity of the bar, the pixel values are expected to remain unchanged within the Doppler-region between the two frames. Therefore, the nonzero region on the difference image indicates the location of the bar. The plot of this location across the frames reveals a zigzag curve and is plotted in Fig. 1(c); the abrupt changes represent the time when the bar reaches the right boundary of the Doppler-region and restarts from the left. The final reconstruction step was to extract the updated segments from the frames and splice them to generate a long Doppler strip with no duplicate or missing data. An example segment of a strip, spanning seven cardiac cycles, is shown in Fig. 1(d).

D. Image Analysis

Doppler images often contain speckle noise in the background and aliasing may also be present. In order to detect the velocity profile from the Doppler tracing, an objective thresholding technique is applied. To this end, the RGB Doppler image in the Doppler-region is converted to gray-scale intensity image by forming a weighted sum of the R, G, and B components $(0.299 \times \mathrm{R} + 0.587 \times \mathrm{G} + 0.114 \times \mathrm{B})$. A binary image is then defined where its pixel values p_{i} for threshold value of P are obtained by

$$p_i = \begin{cases} 0, & I_i < P \\ 1, & I_i \geq P \end{cases}, \quad S_p = \sum_{i=1}^{n} p_i \tag{1}$$

where I_i is the intensity value of pixel I in the gray-scale image. The histogram-index S_p is then the sum of n thresholded pixels

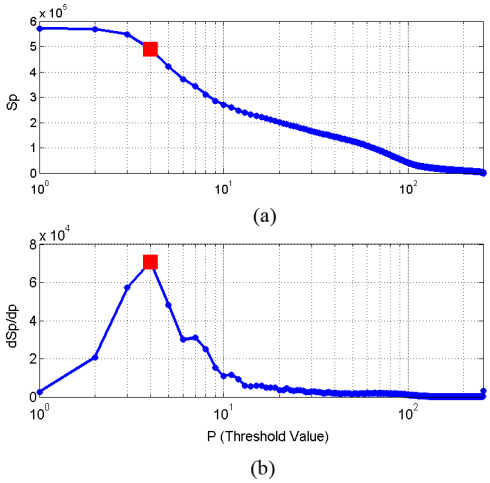

Fig. 2. Objective thresholding: histogram-index S_p of a binary Doppler image (a) and its gradient $\mathrm{d}S_p/\mathrm{d}p$ (b) showing the level of optimum threshold value P (square marker).

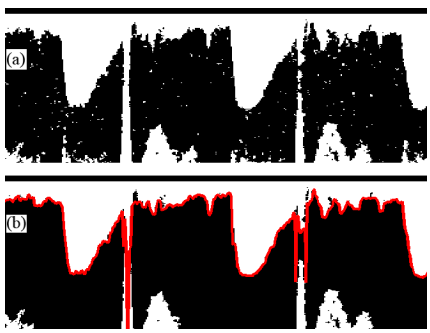

Fig. 3. An example binary Doppler image: (a) noisy/spurious image, (b) same image after noise removal together with superimposed automatically detected initial velocity profile.

in the resulting binary image. Fig. 2(a) shows a typical plot of S_p for different threshold values P ranging from minimum to maximum intensity values in the gray-scale image. An 80% threshold value of the intersection of the abscissa and the tangential line at the steepest gradient of the histogram has been suggested to separate the foreground from the background [23]. However, the position of the steepest gradient of $\mathrm{d}S_\mathrm{p}/\mathrm{d}p$ as shown in Fig. 2(b) deemed to be a good compromise for removing the noise while retaining the Doppler data. Therefore, the optimum threshold is considered as the value for which the largest number of pixels turn from 1 to 0. This value represents the background pixels values in the gray-scale Doppler image. The binary image for the optimum threshold is then adopted for further analysis.

In most cases, the binary image contains an element of noise which manifests as small spurious areas spread throughout the image [Fig. 3(a)]. These are clusters of background pixels that have a higher intensity value than the selected threshold. Adopting a higher value of threshold would remove such noise, but it would also remove segments of the desired velocity envelope and may lead to the underestimation of blood flow. Therefore, in order to filter out the small noisy clusters, connected areas that have fewer than a predefined number of the pixels are removed; this number was empirically selected as 500 pixels.

The maximum velocity profile is then extracted from the resulting filtered image by using the biggest-gap method [4]. This is done by sweeping the image from left to right. Each column of the image represents a vector containing black and white pixels. The gap is defined as a cluster of consecutive black pixels, and the pixel at the beginning of the largest gap from top is selected as one point on the velocity profile [superimposed curve in Fig. 3(b)]. This method allows isolating the desired Doppler envelope from the aliased signal. Nevertheless, there still could be noisy spikes on the extracted profile.

E. ECG-Free Cardiac Timing

Since our automated technique requires only an estimate of the mean cardiac cycle length, and not an exact detection of the QRS complex, we use the Doppler data itself to determine the mean cardiac cycle length. A segment of an example velocity profile, obtained from the long Doppler strip spanning multiple cardiac cycles, is shown in Fig. 4(a). This profile is typically noisy and has some sharp spikes due to the aliasing; also evident in Fig. 3(b).

In order to verify the results of cardiac cycle length estimation, ECG data was collected simultaneously from a subgroup of patients, with heart rates of 60–96 bpm (patients 1–5 in Table IV in results section), using the standard echo machine. The ECG trace, synchronized with the Doppler data, was extracted from the image based on the color of the trace and is plotted in Fig. 4(b).

Autocorrelation, which measures the similarity of a signal as a function of time lag between them and can be used for rate estimation [24], has been suggested for event detection on ECG recordings [25]. Here, the same concept was applied to Doppler velocity profile and it reveals distinct local maxima which decrease in amplitude as the time lag increases; plotted in Fig. 4(c). Here the time lag is measured in pixels and converted to seconds. The first peak point has an amplitude of one with zero time lag. The second, indicates the location of cardiac cycle length where, for the time lag of 0.85 s (164 pixels) in the example shown, the profile has the second highest degree of similarity with itself.

The cardiac cycle lengths estimated using the ECG signal were in good agreement with those obtained from the Doppler velocity profile where a maximum relative error of 4.83% was observed for all recordings with ECG, as provided in Table I. However, all the results reported in this study used only the Doppler data for cardiac timing information.

F. Parameter Extraction

In order to filter out the high frequency noise, a low-pass first-order Butterworth digital filter is applied to the initial velocity profile. The cut-off frequency is estimated from the cardiac cycle length in pixels computed previously; any frequency 10 times higher than the fundamental frequency of the heart motion is filtered out. This ratio was selected empirically as a trade-off between noise removal and maintaining the desired features. The resulting processed profile is relatively smoother and is plotted in Fig. 5, superimposed on the original one. The

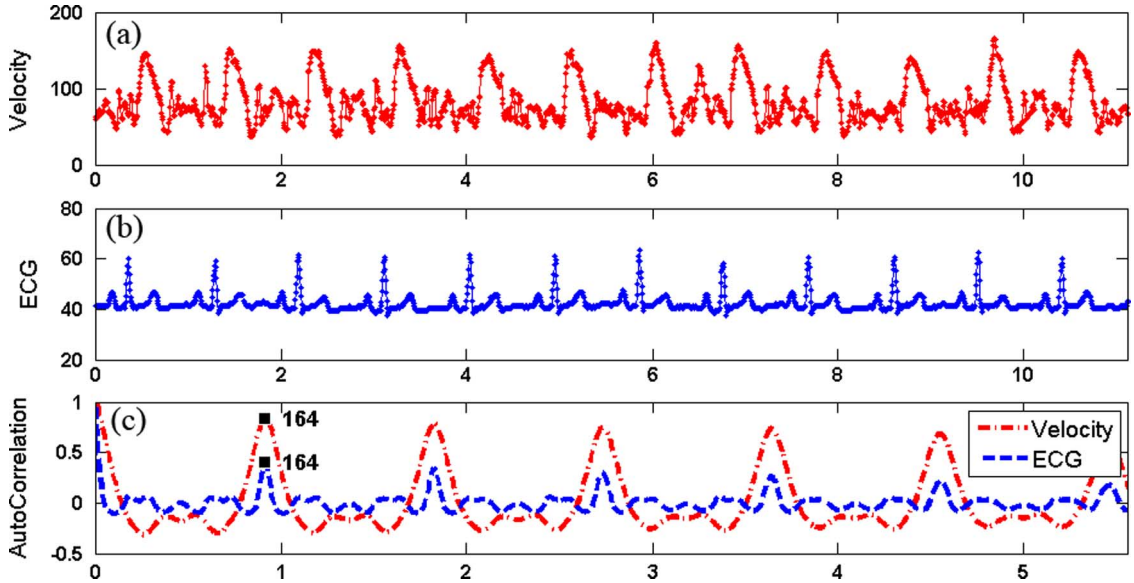

Fig. 4. Estimation of cardiac cycle length (horizontal axis is in seconds for all three panels): (a) a segment of initial crude Doppler velocity profile [inverted with respect to curve in Fig. 3(b)], (b) corresponding ECG signal also extracted from Doppler image (used in this study for validation purposes only), (c) auto-correlation analysis of Doppler and ECG traces where the second peak point is considered as an estimate of cardiac length $(0.85\text{s} \sim 164 \text{ pixels})$.

TABLE I
CARDIAC CYCLE LENGTH ESTIMATED FROM ECG SIGNAL (L-ECG) AND DOPPLER VELOCITY PROFILE (L-DOPPLER) USING AUTOCORRELATION

Patient	L-ECG (s)	L-Doppler (s)	Error (%)
1	0.69	0.67	2.98
2	0.59	0.62	4.83
3	0.84	0.84	0
4	1.00	0.99	0.01
5	0.71	0.72	1.39

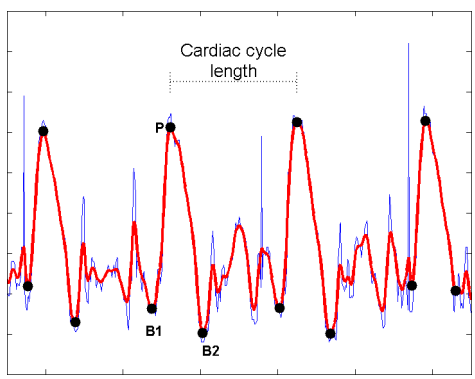

Fig. 5. Initial noisy velocity profile (thin curve) and the processed profile after application of a low-pass filter (thick curve). The location of fiducial points that are used for curve fitting is also shown.

low-pass filter also suppresses the high-amplitude outliers and artefacts in the velocity profile which is crucial for the isolation of the individual cardiac cycles.

In the next processing step, peak points on the smoothed velocity curve are identified by imposing the constraint that the distance between two consecutive peaks should not be smaller than 80% of the cardiac cycle length. This ensured that high-amplitude artefacts, still present after the filtering process, are not selected as genuine peak points. The location of detected peak points are shown as black dots.

Based on heuristic properties of the LVOT flow envelope, the velocity for each single cardiac cycle was considered to start at the base-point B1, reach peak P, and end at B2. In order to detect the location of base-points, the first derivative of the velocity curve is calculated and the immediate local minima on both sides of each peak point are selected.

In order to obtain the final automated LVOT traces, a third-order Gaussian model [26] was fitted to the velocity profile as

$$f(t) = \sum_{n=1}^{N} a_n \cdot \exp\left(-\left(\frac{t-b_n}{c_n}\right)^2\right), \quad N = 3. \qquad (2)$$

The model is extended to beyond the fiducial points B1 and B2 to reach the zero-velocity horizontal axis. The curve fitting is carried out for each individual heartbeat. The peak velocity and VTI value are calculated from the model and the scaling ratios are used to convert pixel units to cm/s and cm.

G. Manual Tracing

Three accredited and experienced cardiology experts manually traced the Doppler flow envelopes in triplicate for 54 images (three from each of the 18 patients). Each beat could therefore undergo up to nine manual measurements. There were 398 candidate Doppler beats and, therefore, up to 3582 manual measurements. Where an operator judged a beat to be of low quality, they declared it invalid and did not make a measurement. However, since each operator viewed each Doppler strip three times, blinded to each other and their own previous measurements, there were beats which were measured on one or two viewings only by each expert.

We developed a custom-made program which closely replicated the interface of echo hardware. Operators manually traced the Doppler images using a track-ball. The experts made their measurements in one or more sessions at their convenience and documented the total time taken. The 54 images were renamed

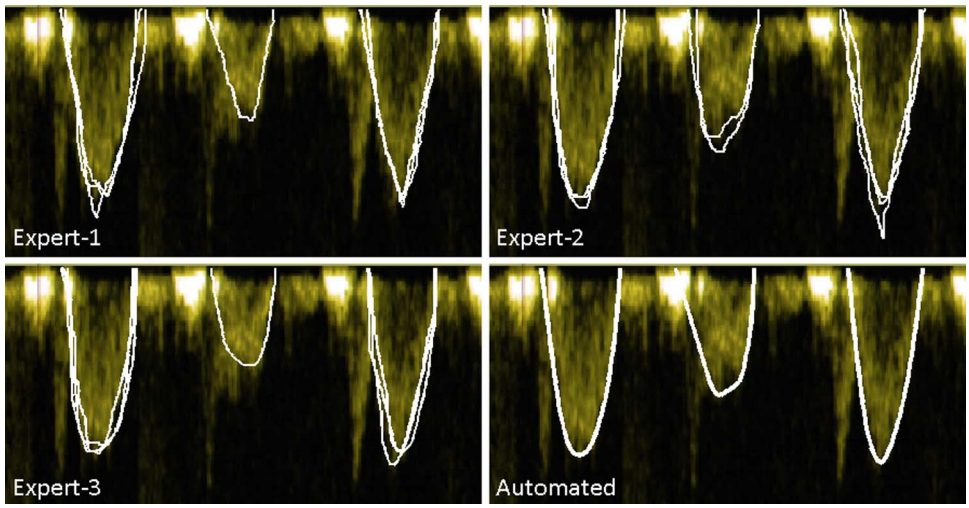

Fig. 6. Example Doppler flow velocity curves showing manual traces for all three experts and the automated traces for the same beats.

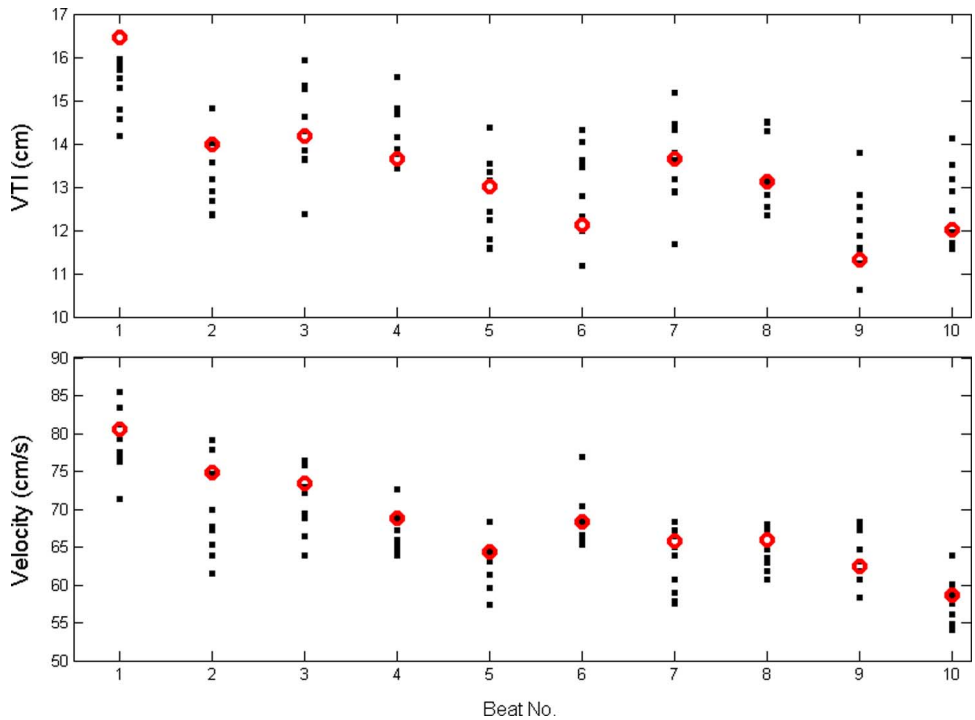

Fig. 7. An example of manual and automated results for VTI (top) and peak velocity (bottom) for the first 10 beats in a Doppler strip. Each of nine manual measurement is marked with a square symbol. The automated measurements are shown as large circles.

and provided to the operators in a random order, and no immediate numerical result was shown after tracing each beat. This way, and given the large number of beats, we made sure that the operators were blinded from their own previous measurements.

H. Data Analysis

In order to quantify the agreement between the two methods, linear regression and Bland–Altman analyses were performed. For each linear regression, the coefficient of determination (R^2) was computed. For the Bland–Altman analysis, bias (mean of the signed differences) and standard deviation (SD) were calculated where the confidence interval was defined as $\pm 2\mathrm{SD}$. The code development for data collection and data analysis was done using $\mathrm{C}++$ and MATLAB programing languages, respectively. All automated computations in this study were conducted using an Intel Xeon E5630 CPU, with an internal clock frequency of 2.53 GHz.

III. Results

Fig. 6 illustrates example velocity curves where the manual and automated traces are superimposed on the Doppler flow envelopes across three consecutive beats. The time used for the manual and automated analysis for each individual beat was $\sim 12\mathrm{s}$ and $\sim 0.5\mathrm{s}$, respectively. For two of the beats, there are three replicate manual traces for each of the experts. For the middle beat, however, experts 1 and 3 had only one replicate

TABLE II
LINEAR REGRESSION AND BLAND-ALTMAN ANALYSES

	Linear Regression				Bland-Altman			
	Peak velocity		VTI		Peak velocity		VTI	
	R^2	*m*, *b*	R^2	*m*, *b*	Bias (cm/s)	SD (cm/s)	Bias (cm)	SD (cm)
Expert 1								
V1	0.96	0.97, 3.38	0.90	0.86, 2.02	-0.35	3.78	0.98	1.53
V2	0.97	1.03, -1.94	0.93	0.97, -0.05	-0.65	3.83	0.66	1.26
V3	0.97	0.98, 0.41	0.93	0.95, 0.46	1.39	3.74	0.54	1.28
Expert 2								
V1	0.96	1.06, -4.22	0.92	1.02, 0.72	-2.26	4.35	-1.17	1.45
V2	0.93	0.98, 4.67	0.93	1.02, 0.76	-2.86	5.18	-1.11	1.30
V3	0.95	1.08, -1.79	0.94	1.04, 0.85	-6.11	5.01	-1.70	1.29
Expert 3								
V1	0.98	0.99, -2.25	0.96	1.09, -2.42	3.25	3.13	0.59	1.15
V2	0.98	0.94, 1.82	0.95	1.01, -0.99	4.05	3.19	0.87	1.14
V3	0.98	0.97, -0.07	0.96	1.05, -1.35	3.55	2.96	0.35	1.05
Average Exp.	**0.96**		**0.94**			**3.91**		**1.27**
Automated	**0.98**	**0.96, 3.23**	**0.94**	**1.09, -1.62**	**1.16**	**3.07**	**-0.24**	**1.39**

Linear regression (R^2: coefficient of determination, fitted line y = *mx*+*b*) and Bland-Altman (SD: standard deviation).
V1 to V3 represent viewings 1 to 3 for manual measurements. Average Exp. is the average performance of all experts (i.e. mean of 9 values above it).

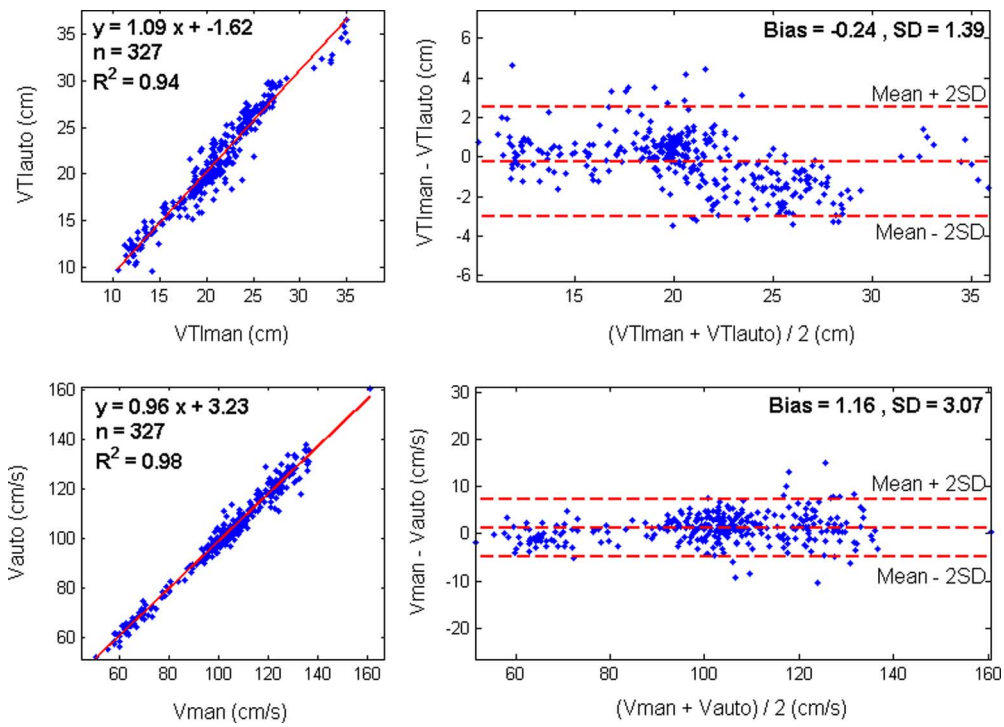

Fig. 8. Linear regression (left) and Bland–Altman (right) analyses between automated and manual measurements for VTI (top) and peak velocity (bottom) for all retained heartbeats in 18 patients; VTIman and Vman are the average of nine manual measurements for VTI and peak velocity, respectively, where x and y axes have the same scale in each panel.

and expert 2 had two replicates. The missing traces are because each expert, on one or two viewings, considered this beat as invalid. In the main analysis, we have used only those beats for which all three experts considered valid on all three occasions. This comprised a total number of 327 heartbeats retained for 18 patients.

A. Beat-by-Beat Comparison

The peak velocity and VTI values were extracted from each manual and automated LVOT curve. In order to compare the manual and automated results, scatter plots of the measurements for a patient are shown in Fig. 7 where only the first 10 beats are shown for the sake of clarity. Scatter plots were chosen over the mean ±SD because it is important to show the distribution of the individual measurements, as we have previously described [27]. We assumed an automated measurement to be acceptable

TABLE III
AVERAGE RESULTS FOR INDIVIDUAL OBSERVERS AND AUTOMATED SYSTEM

	Peak velocity (cm/s)			VTI (cm)		
	Mean	SD	CV%	Mean	SD	CV%
Expert 1	102.17	20.15	19.72	19.99	4.55	22.76
Expert 2	106.04	21.35	20.13	22.04	5.01	22.73
Expert 3	98.68	19.47	19.73	20.11	5.04	25.06
Automated	101.14	19.31	19.09	20.95	5.29	25.26

SD: standard deviation, CV: coefficient of variation

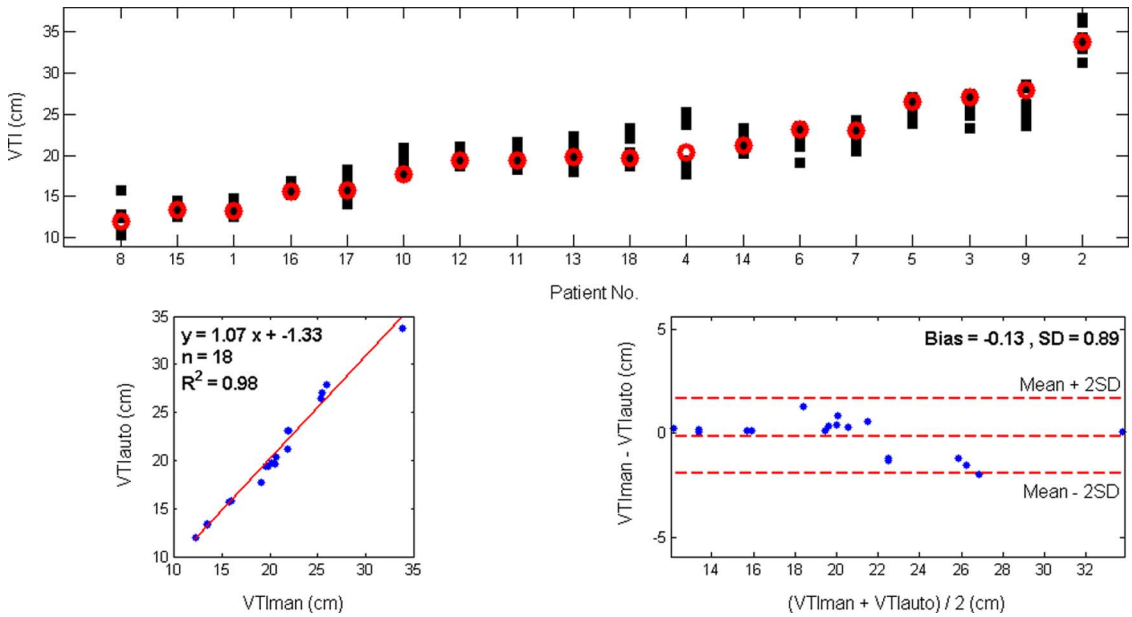

Fig. 9. Patient-by-patient comparisons. Top: plot of VTI measurements as in Fig. 7, but for average values for each patient. Horizontal axis is sorted in ascending order for the average of nine manual measurements. Bottom: linear regression (left) and Bland–Altman (right) analyses; VTIman is the average of nine manual measurements.

when it was within the range of manual estimates. In the example shown, the automated VTI values agree with the manual measurements closely. For beat 1, however, the automated value fell outside the manual range. As for the peak velocity, all automated values lay within the range of manual values. Each of the 327 heartbeats in all 18 patients had 10 measurements; nine manual and one automatic. The automated measurements laid outside the manual measurements for 9.5% (31/327) for VTI and 3.9% (13/327) for peak velocity of heartbeats.

Table II summarizes the results of statistical analyses where the comparison is made between each of the 10 individual measurements episodes and the reference measurement (average of all nine manual measurements). The average coefficient of determination (linear regression) and average standard deviation of differences (Bland–Altman) between a single episode of manual measurement and the reference measurement was $R^2 = 0.96$, SD $= 3.91$ cm/s for peak velocity, and $R^2 = 0.94$, SD $= 1.27$ cm for VTI. Similar results was obtained by comparing the automated measurement with reference measurement ($R^2 = 0.98$, SD $= 3.07$ cm/s for peak velocity, and $R^2 = 0.94$, SD $= 1.39$ cm for VTI). The graphic results of the comparisons for the automated measurements are shown in Fig. 8.

Table III provides an overview of the performance of each expert where the parameters are calculated from the pool of three viewings for each expert (3viewings $\times$ 327 beats). For the automated measurement, 327 beats were used.

B. Patient by Patient Comparison

In addition to beat-by-beat comparisons, we examined the agreement between the two methods in a patient-by-patient fashion by computing the average value across all beats for each patient. The results for the VTI values are depicted in Fig. 9. It is evident that the variability between the experts and between three replicates for each expert is relatively small compared to the variability between the patients. The automated measurements are very close to the reference (average) manual values. None of automated VTI values fell outside the manual range.

The good patient-by-patient agreement between the two methods may be due to averaging multiple cardiac cycles in a Doppler strip that can potentially reduce the effect of potential outliers, for which the discrepancy between different experts and between average manual and automated values may be the most noteworthy. Linear regression and Bland–Altman analyses for VTI measurements were also performed and indicate a very good agreement between the two methods (Fig. 9). Similar results were obtained for the peak velocities; $y = 0.94x + 4.81$, $R^2 = 0.99$ for linear regression, and bias $= 1.41$ cm/s, SD $= 2.44$ cm/s for Bland–Altman.

C. Implication of Beats Being Judged Low Quality

As stated previously, while performing the manual tracing, experts were asked to decide whether each individual beat was of acceptable quality, and to discard the low quality beats that they would not use in clinical practice. In order to investigate the effect of this beat selection, an overlapping samples Student t-test (95% confidence level) was used to compare the two groups of the automated VTI values. The first group comprised all the heartbeats present in each Doppler strip, and the second group was only those beats retained by all experts on all three viewings. The dropped beats in the second group were considered as missing data in the t-test and the results for each individual patient are summarized in Table IV.

For 16 patients, most beats were considered valid by all observers in all viewings. However, those beats that were dropped by experts on some viewings did tend to have smaller values when analyzed by the automatic system than those beats considered valid. This supports the hypothesis that the experts were

TABLE IV
OVERLAPPING SAMPLES STUDENT T-TEST

Pt.	Heart rhythm	HR	All beats			Retained			Dropped			t-test	
			N	Mean	SD	N	Mean	SD	N	Mean	SD	t	*p*-value
1	NSR	89	28	67	14.1	18	75.5	9.6	10	51.7	4.1	-2.25	0.03
2	AF	96	11	122.3	8	10	123.7	6.8	1	108.3	0	-0.43	0.67
3	NSR	71	24	110.4	6.8	24							
4	NSR	60	17	109.4	12.9	11	116.2	3.4	6	97	15.1	-1.69	0.1
5	NSR	83	22	131	7.8	22							
6	NSR	74	25	115.6	7.9	25							
7	NSR	77	26	102.4	5.3	26							
8	NSR	79	21	65.4	2.9	16	65.9	3	5	63.9	2.4	-0.49	0.63
9	NSR	66	22	122.1	4.4	22							
10	AF	89	24	93	6.4	21	94.5	5.4	3	83.1	3.7	-0.79	0.43
11	NSR	72	22	94.5	4.7	18	95.8	3.2	4	88.9	6.9	-0.96	0.35
12	NSR	74	24	100.5	3.8	24							
13	NSR	59	19	101.3	7.1	13	104.8	3.3	6	93.6	7.1	-1.68	0.1
14	NSR	71	24	98.9	4.8	19	98.6	5.3	5	100.3	1.7	0.23	0.82
15	NSR	80	15	65.9	9.6	10	68.3	6.5	5	61.1	13.7	-0.69	0.5
16	AF	71	16	62.3	5.4	15	62.8	5.2	1	54.2	0	-0.28	0.78
17	NSR	93	27	86.2	18.3	7	105	13.7	20	79.7	15	-2.52	0.02
18	NSR	94	31	102	3.6	26	102.3	3.5	5	100.7	4.2	-0.27	0.79

Student t-test for VTI values (in cm) between all beats and beats retained. Dropped beats were considered as missing data. Confidence level was 95%.
N: number of beats, SD: standard deviation in cm, t: t-statistic value, NSR: normal sinus rhythm, AF: atrial fibrillation, Pt.: patient no, HR: estimated heart rate in bpm from cardiac cycle length

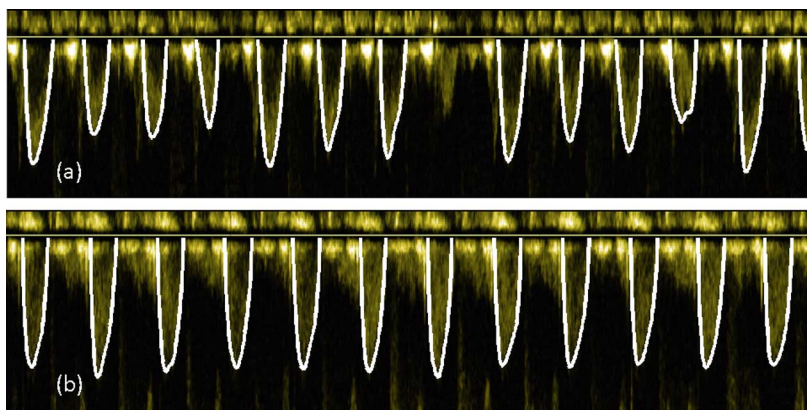

Fig. 10. Two segments of Doppler strips with superimposed automated traces, for (a) patient 1 and (b) patient 3.

choosing to identify traces as invalid that were smaller than surrounding beats.

For two patients, namely patients 1 and 17, there were significant differences in VTI values between the two groups. Not surprisingly, the percentage of dropped beats for these patients was the largest; 36% and 74% for patients 1 and 17, respectively. Fig. 10(a) shows a segment of the Doppler strip together with the superimposed automated traces for patient 1. Large beat-to-beat variations are evident that could be either due to genuine physiological variations or poor quality image. A similar segment for patient 3, with no dropped beats, is shown in Fig. 10(b) where the Doppler envelopes exhibit a steadier pattern.

IV. DISCUSSION

In this work, the feasibility of using an automated method for detecting LVOT Doppler envelopes was demonstrated. Long strips of Doppler images were obtained by a vendor-independent method and analyzed using image processing techniques. Exploiting Doppler images, spanning multiple cardiac cycles, allowed an estimation of the cardiac cycle length to be obtained, without the need for the ECG signal. The performance of the proposed technology was evaluated by comparing the automated VTI and peak velocity values with the consensus of nine manual counterparts, obtained from three expert cardiologists. The results revealed that almost all automated values were within the range of expert measurements. Statistical analyses of linear regression and Bland–Altman showed a good agreement between the two methods. The time-consuming process of drawing around the Doppler traces was reduced by 24-fold using the automated system.

A. Study Limitations

One difference between the human experts and the automated system in our study was that human experts, drawing upon their experience, discarded certain beats as unreliable due to low quality. The results in Table IV showed that the discarded beats tended to be somewhat smaller and this affected the overall average values reported for two patients. Therefore, the proposed automated system should be further improved to equip the technology with means of eliminating low quality data. A comprehensive study, involving a large population of patients with different heart conditions, will ultimately be

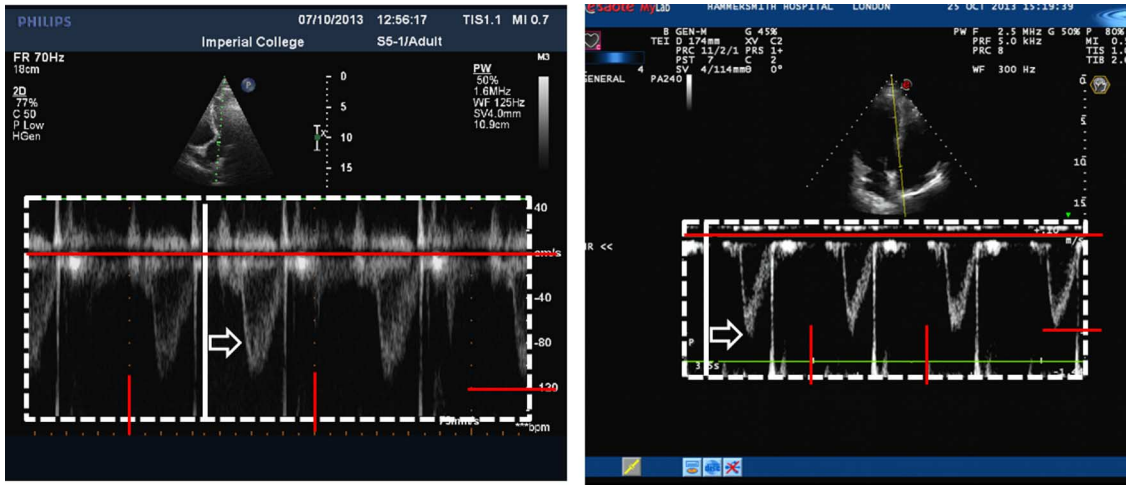

Fig. 11. Collection of snapshots of still-frames and automated extraction of Doppler-region, zero-velocity axis, scaling ratios, and updating sliding bar for Philips (left) and Esaote (right) ultrasound scanners.

required to develop future technology with the heuristics that discriminate genuine physiological variation from those caused by the poor image quality.

This study analyzed Doppler envelopes for LVOT traces which are uniphasic. Dealing with the Mitral flow, which is biphasic, would need a more sophisticated algorithm for detection and quantification of Doppler velocity and envelope. For patients without Mitral valve disease, a simple construct might be to fit the profile to separate peaks. The first peak, the E-wave, is asymmetrical with a tendency for the decline to be slower than the rise. For patients with some degree of Mitral stenosis, it might be necessary to represent the period between the two peaks with an additional nonzero plateau. Similar consideration are required for the Tricuspid valve whose flow is also biphasic. If this approach was to be used with regurgitation, the characteristic pattern of progressive decline would need to be modelled in a different way.

The optimum value for contrast thresholding was empirically selected using the steepest gradient method, and deemed to be the optimum value for segmenting the Doppler images used in this study, obtained from three ultrasound machines. However, a larger dataset of images, from different ultrasound machines, with various levels of signal to noise ratio and contrast to noise ratio should be used to examine the robustness of the thresholding method adopted in this study.

The autocorrelation method was used to estimate the cardiac cycle length from Doppler velocity profile and its performance was examined against the ECG signal for five patients. However, this was an initial pilot study and more complete developments will be required. In atrial fibrillation (AF), cardiac cycle length varies which makes it more difficult to correctly identify the cycles from the Doppler traces alone. In our three AF patients studied here, the variation in cycle length was not so severe as to prevent detection of the individual heartbeats. However, other cases may occur where cycle rate changes by a factor of two or more between successive beats, and more testing is required in this regard. Even though our method relies only on obtaining an approximate value for the cardiac cycle length, and accommodates a substantial degree of heart rate variability, there may turn out to be some patients where the ECG-free algorithm might be not as successful as one using an ECG signal.

B. Clinical and Research Implications

On most commercial echo machines, long Doppler strips are not easily available to end-users and only snapshots of still-frames can be exported and analyzed. One aim in our study was to develop a low-cost and vendor-independent technology that could be applied to majority of the echo hardware with a standard graphics output. Most ultrasound scanners from all manufactures are equipped with an external video output. With this automated technology in place, a variety of different studies could be conducted using any type of echo machine.

One major source of variability in Doppler measurements is human variability [28]. This may be introduced at two stages: performing the scan, and extracting the clinical measurements from the collected data. While the inter- and intra-observer variability in parameter extraction could potentially be eliminated using an automated system, the effects of subtle differences in probe positioning (angle of insonation) and choices of placement of sample volume would still be present. One way to handle this using the automated system is to take the average of measurements in several positions that all appear clinically valid, but incorporate small differences which are not noticed clinically. In a future study, we plan to investigate the effect of averaging the measurements at subtly different positions in reducing the variations.

Lack of experience could also be a contributing factor to the variability between different operators. Our technology may have a role in improving the quality of measurements made by operators who have not undergone traditional exhaustive training in echocardiography.

It has been suggested that respiration artefacts limit the use of long sequences for measuring Doppler velocities and, therefore, including only a few beats with breath held has been advocated. However, an average of many heartbeats during normal free-breathing may be more representative. The effect of respiration on Doppler measurements can be examined by acquiring long strips with free-breathing and short breath-hold strips.

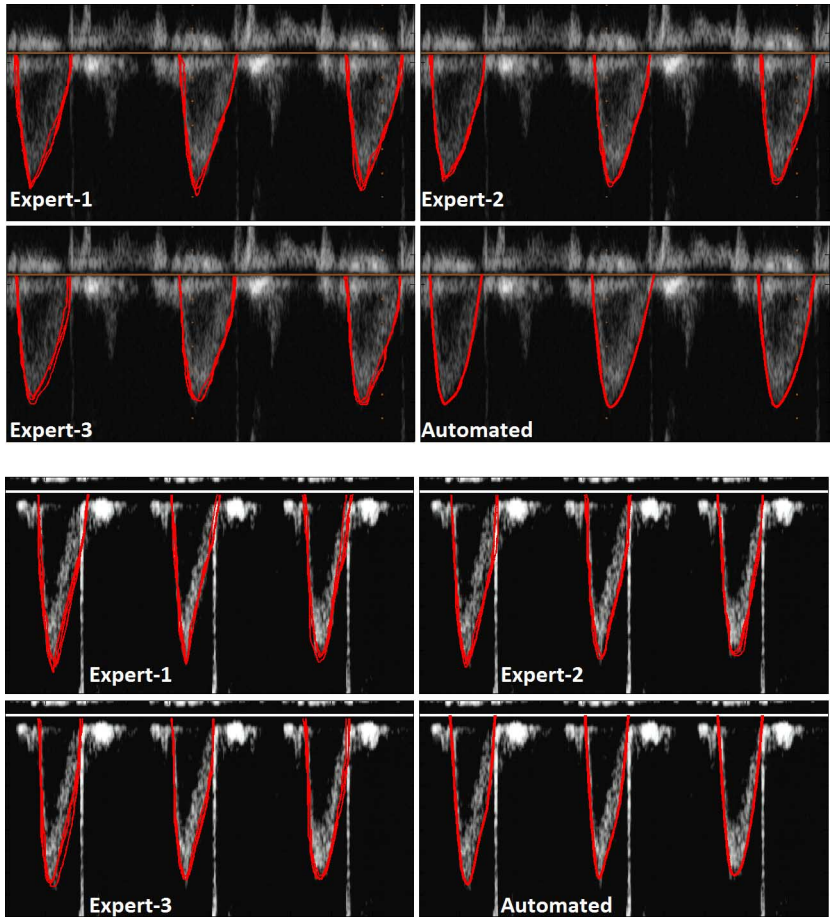

Fig. 12. Example Doppler flow velocity curves showing manual traces for all three experts and the automated traces for the same heartbeats collected from Philips (top) and Esaote (bottom) ultrasound scanners.

Although dedicated Doppler systems (e.g., USCOM, Sydney, Australia) provide some means of automated measurements for Doppler traces, the idea behind our proposed system is to enhance the ability of current physiological researchers to make measurements with conventional and existing equipment (which provides imaging information to confirm localization of the beam) in a manner with which they will be familiar, without the need for deployment of the new hardware.

APPENDIX

Fig. 11 illustrates the graphics results for LVOT Doppler data collected from Philips (frames with resolution of 1050×1680) and Esaote (frames with resolution of 768×1024) ultrasound scanners. Fig. 12 shows the manual and automated traces for the two scanners.

REFERENCES

[1] A. Vahanian, O. Alfieri, F. Andreotti, M. J. Antunes, G. Barón-Esquivias, H. Baumgartner, M. A. Borger, T. P. Carrel, M. D. Bonis, and A. Evangelista, "Guidelines on the management of valvular heart disease (version 2012) the joint task force on the management of valvular heart disease of the European society of cardiology (ESC) and the European association for cardio-thoracic surgery (EACTS)," *Eur. Heart J.*, vol. 33, pp. 2451–2496, 2012.

[2] J. A. Finegold, C. H. Manisty, F. Cecaro, N. Sutaria, J. Mayet, and D. P. Francis, "Choosing between velocity-time-integral ratio and peak velocity ratio for calculation of the dimensionless index (or aortic valve area) in serial follow-up of aortic stenosis," *Int. J. Cardiol.*, vol. 167, pp. 1524–1531, 2013.

[3] J. Tschirren, R. M. Lauer, and M. Sonka, "Automated analysis of Doppler ultrasound velocity flow diagrams," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1422–1425, Dec. 2001.

[4] H. Greenspan, O. Shechner, M. Scheinowitz, and M. S. Feinberg, "Doppler echocardiography flow-velocity image analysis for patients with atrial fibrillation," *Ultrasound Med. Biol.*, vol. 31, pp. 1031–1040, 2005.

[5] O. Shechner, M. Scheinowitz, M. Feinberg, and H. Greenspan, "Automated method for Doppler echocardiography image analysis," *Proc. 23rd IEEE Convent. Elect. Electron. Eng. Israel*, pp. 177–180, 2004.

[6] P. Shechner, M. Sheinovitz, M. Feinberg, and H. Greenspan, "Image analysis of Doppler echocardiography for patients with atrial fibrillation," in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano to Macro*, 2004, pp. 488–491.

[7] P. Turaga, D. Beymer, F. Wang, A. Amir, H. Greenspan, and K. Pohl, "Shape-based similarity retrieval of Doppler images for clinical decision support," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 855–862.

[8] V. Magagnin, L. Delfino, S. Cerutti, M. Turiel, and E. G. Caiani, "Nearly automated analysis of coronary Doppler flow velocity from transthoracic ultrasound images: Validation with manual tracings," *Med. Biol. Eng. Comput.*, vol. 45, pp. 483–493, 2007.

[9] S. K. Zhou, F. Guo, J. Park, G. Carneiro, J. Jackson, M. Brendel, C. Simopoulos, J. Otsuki, and D. Comaniciu, "A probabilistic, hierarchical, and discriminant framework for rapid and accurate detection of deformable anatomic structure," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[10] J. Park, S. K. Zhou, J. Jackson, and D. Comaniciu, "Automatic Mitral valve inflow measurements from Doppler echocardiography," in *Medical Image Computing and Computer Assisted Intervention-MICCAI* . New York: Springer, 2008, pp. 983–990.

[11] E. Gaillard, L. Kadem, M. Clavel, P. Pibarot, and L. Durand, "Optimization of Doppler echocardiographic velocity measurements using an automatic contour detection method," *Ultrasound Med. Biol.*, vol. 36, pp. 1513–1524, 2010.

[12] C. H. Manisty, A. Al-Hussaini, B. Unsworth, R. Baruah, P. A. Pabari, J. Mayet, A. D. Hughes, Z. I. Whinnett, and D. P. Francis, "The acute effects of changes to AV delay on BP and stroke volume potential implications for design of pacemaker optimization protocols," *Circulation: Arrhythmia Electrophysiol.*, vol. 5, pp. 122–130, 2012.

[13] P. A. Pabari, K. Willson, B. Stegemann, I. E. van Geldorp, A. Kyriacou, M. Moraldo, J. Mayet, A. D. Hughes, and D. P. Francis, "When is an optimization not an optimization? evaluation of clinical implications of information content (signal-to-noise ratio) in optimization of cardiac resynchronization therapy, and how to measure and maximize it," *Heart Fail. Rev.*, vol. 16, pp. 277–290, 2011.

[14] D. P. Francis, "How to reliably deliver narrow individual-patient error bars for optimization of pacemaker AV or VV delay using a pick-the-highest" strategy with haemodynamic measurements," *Int. J. Cardiol.*, vol. 163, pp. 221–225, 2013.

[15] S. Sohaib, Z. I. Whinnett, K. A. Ellenbogen, C. Stellbrink, T. A. Quinn, M. D. Bogaard, P. Bordachar, B. M. v. Gelder, I. E. van Geldorp, and C. Linde, "Cardiac resynchronisation therapy optimisation strategies: Systematic classification, detailed analysis, minimum standards and a roadmap for development and testing," *Int. J. Cardiol.*, vol. 170, pp. 118–131, 2013.

[16] C. Kupfahl, M. Honold, G. Meinhardt, H. Vogelsberg, A. Wagner, H. Mahrholdt, and U. Sechtem, "Evaluation of aortic stenosis by cardiovascular magnetic resonance imaging: Comparison with established routine clinical techniques," *Heart*, vol. 90, pp. 893–901, 2004.

[17] A. Margulescu, M. Cinteza, and D. Vinereanu, "Reproducibility in echocardiography: Clinical significance, assessment, and comparison with other imaging methods," *Mædica*, vol. 1, pp. 29–36, 2006.

[18] M. Galderisi, A. Santoro, M. Versiero, V. S. Lomoriello, R. Esposito, R. Raia, F. Farina, P. L. Schiattarella, M. Bonito, and M. Olibet, "Improved cardiovascular diagnostic accuracy by pocket size imaging device in non-cardiologic outpatients: the nausica (Naples ultrasound stethoscope in Cardiology) study," *Cardiovasc. Ultrasound*, vol. 8, pp. 1506–1513, 2010.

[19] S. Brekke, E. Tegnander, H. Torp, and S. Eik-Nes, "Tissue Doppler gated (TDOG) dynamic three-dimensional ultrasound imaging of the fetal heart," *Ultrasound Obstetr. Gynecol.*, vol. 24, pp. 192–198, 2004.

[20] P. Gifani, H. Behnam, A. Shalbaf, and Z. A. Sani, "Automatic detection of end-diastole and end-systole from echocardiography images using manifold learning," *Physiol. Meas.*, vol. 31, p. 1091, 2010.

[21] S. A. Aase, S. R. Snare, H. Dalen, A. Støylen, F. Orderud, and H. Torp, "Echocardiography without electrocardiogram," *Eur. J. Echocardiogr.*, vol. 12, pp. 3–10, 2011.

[22] M. A. Quiñones, C. M. Otto, M. Stoddard, A. Waggoner, and W. A. Zoghbi, "Recommendations for quantification of Doppler echocardiography: A report from the Doppler quantification task force of the nomenclature and standards committee of the American Society of Echocardiography," *J. Am. Soc. Echocardiogr.*, vol. 15, pp. 167–184, 2002.

[23] S. Y. Lee and Y. D. Kim, "Sizing of spray particles using image processing technique," *KSME Int. J.*, vol. 18, pp. 879–894, 2004.

[24] J. J. Goldberger and J. Ng, *Practical Signal and Image Processing in Clinical Cardiology*. New York: Springer, 2010.

[25] T. Syeda-Mahmood, D. Beymer, and F. Wang, "Shape-based matching of ECG recordings," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 2012–2018.

[26] P. Richter, Estimating errors in least-squares fitting TDA Progress Rep., 1995.

[27] M. J. Shun-Shin and D. P. Francis, "Why even more clinical research studies may be false: Effect of asymmetrical handling of clinically unexpected values," *PloS One*, vol. 8, p. E65323, 2013.

[28] E. Y. Lui, A. H. Steinman, R. S. Cobbold, and K. W. Johnston, "Human factors as a source of error in peak Doppler velocity measurement," *J. Vasc. Surg.*, vol. 42, pp. 972. e1–972. e10, 2005.