Distinguishing Innocent Murmurs from Murmurs caused by Aortic Stenosis by Recurrence Quantification Analysis

Christer Ahlstrom, Katja Höglund, Peter Hult, Jens Häggström, Clarence Kvart, and Per Ask

Abstract—It is sometimes difficult to differentiate between innocent murmurs and pathological murmurs during auscultation. In these difficult cases, an intelligent stethoscope with decision support abilities would be of great value. In this study, using a dog model, phonocardiographic recordings were obtained from 27 boxer dogs with various degrees of aortic stenosis (AS) severity. As a reference for severity assessment, continuous wave Doppler was used. The data were analyzed with recurrence quantification analysis (RQA) with the aim to find features able to distinguish innocent murmurs from murmurs caused by AS. Four out of eight investigated RQA features showed significant differences between innocent murmurs and pathological murmurs. Using a plain linear discriminant analysis classifier, the best pair of features (recurrence rate and entropy) resulted in a sensitivity of 90% and a specificity of 88%. In conclusion, RQA provide valid features which can be used for differentiation between innocent murmurs and murmurs caused by AS.

Keywords—Bioacoustics, murmur, phonocardiographic signal, recurrence quantification analysis.

I. INTRODUCTION

The relationship between blood volumes, pressures and flows within the heart determines the opening and closing of the heart valves. Normal heart sounds occur during the closure of the valves, while pathological murmurs are produced by turbulent blood flow caused by narrowed or leaking valves (or from the presence of abnormal passages in the heart) [1]. Heart murmurs are common during childhood, between 50 – 80%, but only about 1% of these murmurs are pathological [2]. The clinical standard for assessing valvular heart disease is echocardiography. However, in the primary health care, when deciding who requires special care, auscultation plays an important role. Distinguishing innocent heart murmurs from mild pathological murmurs is a diagnostic challenge, and current murmur assessments as well as echocardiographic evaluations are sometimes inconclusive.

With the advent of digital stethoscopes and the development of portable processing power, we envision an intelligent stethoscope with decision support abilities [3]. Such a tool, able to screen murmurs, would be both time- and cost-saving while relieving many patients from needless anxiety. In this study we used recurrence quantification analysis (RQA) for differentiation between innocent murmurs and murmurs caused by aortic stenosis (AS). The presented work, which is using a dog model, is part of a wider investigation concerning the prognostic value of heart murmur evaluation in young individuals, as well as differentiation between innocent murmurs and murmurs caused by AS in adults [4].

Research on signal processing of the phonocardiographic signal (i.e. recorded sounds from the heart) has been extensive [5]–[6], and several authors have reported tools to screen innocent murmurs from pathological murmurs [7]–[9]. Common for these approaches are time and/or frequency based features which are used as input to an artificial neural network classifier. Spectral features are reasonable since there is an established relationship between the murmur’s frequency content and the severity of the stenosis [10]. However, as stated by Bhatikar et al. [7], the relationship between the features and the target value (the clinical diagnosis) is complex. Advanced artificial neural networks with nonlinear decision boundaries have therefore been applied in the classification task. We suggest that by finding relevant features, the relationship between features and targets can be simplified, thus reducing the risk of misclassification caused by overfitting.

Recent studies indicate that the phonocardiographic signal is nonlinear [11]–[13], thus motivating the use of nonlinear analysis tools. Here RQA is used to capture the nonlinear dynamics of the systolic segment of the phonocardiographic signal. Our hypothesis is that the turbulence-induced murmur signal can be quantified with RQA, especially with the parameters recurrence rate (which should decrease as the
signal becomes more turbulent) and entropy (which should increase by increasing turbulence).

As in the work by Tavel et al. [14], this study is restricted to murmurs induced by aortic stenosis. Our intention is not to replace previously developed classification methods [7]-[9], [14], but rather to complement the previously used features with a new set of features able to capture the nonlinear dynamics of the phonocardiographic signal. More specifically, the aims of this paper are to introduce RQA features to quantify phonocardiographic signals and to prove their usefulness by classifying innocent murmurs from murmurs caused by AS.

II. MATERIALS AND METHODS

The study was approved by the Local Ethical Committee in Uppsala, Sweden. Twenty-seven privately owned boxer dogs were included in the study, 15 females and 12 males, aged (mean ± std) 2.15 ± 2.18 years (range 1-9 years). The inclusion criterion was that an auscultatory heart murmur over the aortic area should be present, whereas other findings indicative of systemic disease or other organ malfunctions on physical examination should be absent. For characterization purposes, all dogs underwent an echocardiographic examination. In order to evaluate heart sounds and murmurs a phonocardiographic examination was performed. Cardiac diseases, congenital as well as acquired, other than AS were excluded by echocardiographic examination. In dogs with innocent murmurs, anemia was excluded by routine hematometry. All examinations were performed by experienced examiners.

A. Measurements

A phonocardiographic examination took place in a quiet room with the dog in a standing position. The examinations were performed using the Welch Allyn Meditron stethoscope (Meditron ASA, Medi-Stim ASA, Oslo, Norway), connected to a computer with the Meditron Analyzer software (Meditron ASA, Medi-Stim ASA, Oslo, Norway), and a Welsh Allyn Meditron stethoscope. The chest piece of the stethoscope was placed over the aortic area, giving the loudest and clearest heart murmur possible. One recording lasted for 10 seconds and was performed on each dog. The signal was digitized at 44.1 kHz with 16 bits per sample using a sound card (Meditron ASA, Medi-Stim ASA, Oslo, Norway). All recordings were stored on the computer.

TABLE I

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of dogs</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Vmax (m/s), mean ± std</td>
<td>1.65 ± 0.09</td>
<td>2.02 ± 0.19</td>
<td>2.82 ± 0.36</td>
<td>4.68 ± 0.57</td>
</tr>
<tr>
<td>Aortic flow velocity, range (m/s)</td>
<td>1.52-1.73</td>
<td>1.84-2.41</td>
<td>2.40-3.20</td>
<td>4.00-5.50</td>
</tr>
<tr>
<td>2D morphological aortic stenosis</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

An echocardiographic examination was performed with the dog in right and left lateral recumbency on the ultrasound examination table. The examinations were performed using a GE Vivid 3 ultrasound machine with a 5 MHz transducer. A complete echocardiographic examination with standardized imaging planes was performed [15], paying special attention to 2D changes indicative of aortic or pulmonic stenosis. The mitral, tricuspid, aortic and pulmonic valves were screened for regurgitation using color flow Doppler. Pulsed-wave Doppler was used for measurement of the pulmonic flow velocity. The peak aortic flow velocity (Vmax) was measured by continuous wave Doppler using the subcostal transducer location recommended by Lehmkuhl and Bonagura [16]. The Vmax parameter was used as a hemodynamic reference to assess AS severity.

B. Group Denomination

The dogs were divided into four classes. Class 1 and 2 had no morphological evidence of AS on 2D echocardiography. Class 1 consisted of dogs with Vmax <1.8 m/s while class 2 had Vmax ≥1.8 m/s. Class 3 and 4 on the other hand had morphological evidence of AS on 2D echocardiography. Class 3 consisted of dogs with Vmax ≤3.2 m/s (mild AS) while class 4 had Vmax >3.2 m/s (moderate-severe AS). The group denomination is based on Kittleson and Kienle [17]. Echocardiographic and auscultatory data are summarized in Table I, while examples of phonocardiographic recordings are shown in Fig. 1.

![Fig. 1 Examples of phonocardiographic signals from dogs with different degrees of aortic stenosis.](image-url)
C. Preprocessing of the Phonocardiographic Signals

All recorded phonocardiographic signals were manually segmented by experienced examiners. Four markers per heart cycle were determined: the beginning of the first heart sound, the end of the first heart sound, the beginning of the second heart sound and the end of the second heart sound. Noisy or corrupted signal segments, determined by visual inspection by KH, were excluded from further studies. The systolic part of the signal, defined as the period ranging from the end of the first heart sound to the beginning of the second heart sound, was used as input for RQA. A 5th order Butterworth high-pass filter with a cut-off frequency of 50 Hz was used to remove trends from the signal. The filtering was performed by zero-phase digital filters, processing the input data in both the forward and reverse directions. The signals were also down sampled to 14.7 kHz to reduce computational complexity. All processing of the phonocardiographic signals were made in MATLAB (The MathWorks, Inc., Natick, MA, USA) and in the cross recurrence plot toolbox [18].

D. Nonlinear Systems and Embedology

\[
x(t+1) = \varphi(x(t)),
\]

where \( \varphi \) is a mapping function such that \( \varphi: \mathbb{M} \rightarrow \mathbb{M} \) and \( M \) is the true state space. In the phonocardiographic signal, only a single scalar measure \( s(t)=h(x(t)), t=1,2,...,N \), is available. If \( s(t) \) is a projection from \( M \), then Takens’ theorem provides a way to undo the projection and reconstruct a dynamically and topologically equal replica of the true state space [19]:

\[
F : M \rightarrow \mathbb{R}^d
\]

\[
x(t) \rightarrow y(t) = F(x(t)) = [s(t), s(t+\tau),...,s(t+(d-1)\tau)]
\]

where \( \tau \) is a delay parameter, \( d \) is the embedding dimension and \( F \) is a map from the true state space to the reconstructed state space. In this study, \( \tau \) was set to 92 based on average mutual information [19] and \( d \) was set to 4 based on Cao’s method [20], see Fig. 2. Since the dynamics of the reconstructed state space contains the same topological information as the original state space, characterization and prediction based on the reconstructed state space is as valid as if it were made in the true state space.

E. Recurrence Quantification Analysis

The state space of a system is often high-dimensional, especially when reconstructed from experimental data where noise tends to inflate the dimension. Its phase portrait can therefore only be visualized by projection into two or three dimensions. This operation does however fold the attractor, and by doing so, destroying its structure. A recurrence plot (RP) is a way to visually investigate the \( d \)-dimensional state space trajectory through a two-dimensional representation [18], [19]. An RP is a symmetric \( N \times N \) matrix that represents the recurrence of states of the system, and a point in coordinate \((i,j)\) indicates that two states \(y(i)\) and \(y(j)\) are close to each other. An RP is defined as:

\[
RP(i, j) = \Theta(\varepsilon - \|y(i) - y(j)\|)
\]

where \( i,j = 1,...,N \), \( \varepsilon \) is a cut-off distance, \( \| \cdot \| \) is the Euclidean norm (any norm could be used) and \( \Theta(\cdot) \) is the Heaviside function. States that are close to each other in the reconstructed state space are represented by black dots in the recurrence plot. An example RP is shown in Fig. 3.

Recurrence quantification analysis is a way to parameterize the RP. Isolated recurrence points occur if states are rare, if they do not persist for any time or if they fluctuate heavily. Diagonal lines occur when a segment of the trajectory runs in parallel with another segment, i.e. when the trajectory visits the same region of the phase space at different times. Vertical (horizontal) lines mark a time length in which a state does not change or changes very slowly. The eight most common RQA-parameters are used in this study [18], [21]:

- Recurrence rate: The percentage of recurrence points (black dots) in the recurrence matrix.
- Determinism: The percentage of the recurrence points...
that form diagonal lines. Diagonal lines are associated with deterministic patterns in the dynamics, hence determinism.

- Mean diagonal line length: The average length of the diagonal lines.
- Maximal diagonal line length: The length of the longest diagonal line. Inversely proportional to the largest Lyapunov exponent which describes how fast trajectories diverge in the reconstructed state space.
- Entropy: The Shannon entropy of the distribution of the diagonal line lengths. Measures the complexity of the signal.
- Laminarity: The percentage of recurrence points which form vertical lines.
- Trapping time: The average length of the vertical lines.
- Maximal vertical line length: The length of the longest vertical line.

In this study, RQA was applied to the systolic period of each heart cycle in each dog. The obtained values were then averaged within each dog resulting in eight RQA feature values per dog.

F. Statistical Analysis

The statistical analysis was made in two steps. The clinically interesting task of separating class 1+2 (dogs without morphological evidence of AS on 2D echocardiography) from class 3+4 (dogs with morphological evidence of AS on 2D echocardiography) was tested using the Wilcoxon rank-sum test. A nonparametric test was used since the assumption of normality could not be verified.

Linear discriminant analysis (LDA) was applied to investigate the separability between class 1+2 and class 3+4 when using the RQA features that were found significant in the previous step [22]. The RQA features that showed significant differences between the groups were analyzed pair wise. Due to the limited study population, a leave one out approach was used to create the training and the test set [23]. This means that all but one dog was used as training data to construct the discriminant functions while the excluded dog was used for validation. This procedure was iterated for all dogs, where a different dog was excluded each time.

III. RESULTS

The number of heart cycles (mean ± std) in the twenty-seven examined boxer dogs were 11 ± 3. According to the Wilcoxon rank-sum test, there were significant differences at the 2% level between dogs from class 1+2 compared to dogs from class 3+4 in four of the eight RQA measures; recurrence rate, entropy, trapping time and maximal vertical line length, see Table II. An example RP is shown in Fig. 3 while box and whisker plots of all eight RQA measures are shown in Fig. 4. Results from the pair wise linear discriminant analysis are shown in Table III.

IV. DISCUSSION

Features obtained via RQA were investigated in this paper. Four of the eight RQA measures showed significant differences between innocent murmurs and murmurs caused by AS at the 2% level.

Recognition of severe AS is not difficult in clinical practice since the murmur generally becomes louder and longer with increasing severity of the obstruction [24]. Mild obstructions, however, cause soft murmurs which are difficult to distinguish from innocent murmurs. This can be seen in Fig. 1. The class denomination in this study is based on Kienle and Kittleson [17], as well as our own clinical experience, considering class 1 as dogs with innocent murmurs and class 2 as dogs in the grey-zone for AS. Both of these classes are however
considered as innocent murmurs due to their limited influence on physiological function. Instead, it is the differentiation between mild AS and dogs without AS which is the most critical in the clinical situation (that is, class 1+2 vs. class 3+4). Dogs with mild AS in our study had auscultatory murmurs between II and IV out of VI, while murmurs in class 1 and 2 varied between I and II out of VI. In dogs with murmur degree II, it is not possible to determine the cause of the murmur by cardiac auscultation alone, and in these dogs the echocardiographic examination is sometimes inconclusive [17]. In the current study, RQA proved useful in separating these difficult groups.

The canine cardiovascular and respiratory systems are similar to the human, which has made the dog a commonly used model in cardiovascular research [25]. In the boxer breed, the prevalence of heart murmurs is high, reportedly between 50 – 80% [17], [26], [27]. A proportion of these murmurs are caused by AS, which is a common heart disease in boxer dogs [26], whereas the underlying cause for the murmur remains undiscovered in other cases. These murmurs are commonly referred to as innocent murmurs. In contrast to humans, canine AS is not of degenerative origin, but a congenital disease affecting the aortic valves or the left ventricular outflow tract due to a subvalvular fibrous stenosis [17]. Another difference is the higher heart rate in the dog, which, in this study, admitted rather short recording sessions (10 seconds) when recording the phonocardiographic signal.

The nonlinear features used in this study are not easy to interpret. When leaving the well known concepts of time and frequency, the obtained features become hard to explain in terms of physiological events. Another complicating issue is the fact that the reconstructed state space is four-dimensional, making it impossible to visualize. Nonetheless, the four significant RQA features may be interpreted as follows. The recurrence rate corresponds with the probability that a specific state will recur, and as the turbulence increases, the probability of recurring states decreases. This is in agreement with Fig. 4a. Entropy reflects the complexity of the deterministic structure in the system. As the turbulence increases, the complexity of the signal increases, see Fig. 4e.

| TABLE III |
|---|---|---|
| Correct classifications (%) | Sensitivity (%) | Specificity (%) |
| Recurrence rate ↔ Entropy | 89 | 90 | 88 |
| Recurrence rate ↔ Trapping time | 85 | 82 | 88 |
| Recurrence rate ↔ Maximal vertical line length | 78 | 72 | 81 |
| Recurrence rate ↔ Trapping time | 85 | 89 | 83 |
| Recurrence rate ↔ Maximal vertical line length | 74 | 75 | 74 |
| Entropy ↔ Maximal vertical line length | 78 | 69 | 86 |

Fig. 4 Box and whisker plots of the eight RQA features showing their distribution in the four defined classes (ranging from innocent murmur to flow murmurs caused by severe aortic stenosis). There is one box for each class and the boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.
Trapping is related with the laminarity time of the system, i.e. how long the system remains in a specific state. This measure should decrease with increasing turbulence, as in Fig 4g. The maximal vertical line length represents the longest segment which remains in the same phase space region over some time. This kind of structure in state space will also decrease with increasing turbulence, see Fig 4h.

In supervised learning, overfitting is likely to occur in cases where (i) learning was performed too long, where (ii) training examples are rare, where (iii) too many feature vectors are used or (iv) a combination thereof. This basically means that many different solutions are consistent with the training examples, but disagree on unseen data. Hence, when presenting new examples to the developed classifier, the predictions will not be reliable. In order to avoid overfitting, it is necessary to use cross-validation to verify the results [23].

Due to the limited amount of data in this study, a leave-one-out methodology was used for cross-validation [23].

In this study, four significant feature vectors were derived. However, these are only used pair wise in the linear discriminant analysis. The reason is to avoid overfitting of the classifier. When using many features in combination with limited training examples, the risk of losing generality is high. Using three of four feature vectors creates a high dimensional feature space which cannot be represented accurately by 26 examples (the last dog is used for cross-validation). This also indicates the finding of importance of representative features. For example, Bhatikar et al. [7] uses 252 features and a nonlinear classifier to achieve a sensitivity of 93% and a specificity of 90% when classifying innocent murmurs from murmurs caused by ventricular septal defect (based on 153 training examples and 88 test examples). In this study, the best feature vectors achieved a sensitivity of 90% and a specificity of 88% when classifying innocent murmurs from murmurs caused by aortic stenosis. The results are comparable to ours, but here we use a plain linear classifier with only two feature vectors.

To conclude, RQA provides valid features which can be used for differentiation between innocent murmurs and murmur caused by AS.

REFERENCES