Noise detection during heart sound recording using periodicity signatures

Noise detection during heart sound recording using periodicity signatures

D Kumar[1], **P Carvalho**[1], **M Antunes**[2], **R P Paiva**[1] **and J Henriques**[1]

[1] Center of Informatics and Systems, University of Coimbra, Coimbra, Portugal
[2] Center of Cardio-thoracic Surgery of the University Hospital of Coimbra, Coimbra, Portugal

E-mail: dinesh@dei.uc.pt

Abstract
Heart sound is a valuable biosignal for diagnosis of a large set of cardiac
diseases. Ambient and physiological noise interference is one of the most
usual and highly probable incidents during heart sound acquisition. It tends
to change the morphological characteristics of heart sound that may carry
important information for heart disease diagnosis. In this paper, we propose a
new method applicable in real time to detect ambient and internal body noises
manifested in heart sound during acquisition. The algorithm is developed on
the basis of the periodic nature of heart sounds and physiologically inspired
criteria. A small segment of uncontaminated heart sound exhibiting periodicity
in time as well as in the time-frequency domain is first detected and applied
as a reference signal in discriminating noise from the sound. The proposed
technique has been tested with a database of heart sounds collected from 71
subjects with several types of heart disease inducing several noises during
recording. The achieved average sensitivity and specificity are 95.88% and
97.56%, respectively.

Keywords: noise detection, heart sound processing, periodicity analysis

1. Introduction

Heart sound is a consequence of turbulent blood flow and vibrating cardiovascular structures,
which propagates through the chest. These vibrations typically result from myocardial and
valvular events that are affected by the function, the hemodynamics and electrical activity
of the muscle. The later have a direct impact on the morphological, spectral and timing
characteristics of the main heart sounds (HSs), which have been found to be highly sensitive
and specific for several important diagnosis tasks ranging from heart valve dysfunction (Durand

and Pibarot 1995, Abbas and Bassam 2009) to the systolic cardiac function (Paiva *et al* 2009, Tavel 1967).

Being an acoustic signal, HS acquisition is prone to the interference of several noise sources that are difficult to avoid, even in well-controlled home or clinical environments. These noise sources include both external ambient noise sources (e.g. ambient music, noises induced by bystanders, etc) and internal body noise sources (e.g. coughing, physiological noises, etc), which interfere with HS in highly complex and unpredictable ways. These noise sources exhibit a very broad range of spectral characteristics overlapping the typical HS frequency range and a broad range of durations and loudness, which might alter prominent diagnosis features in HSs. In this context, detection of inadequate signal acquisition is of primary importance in order to timely alert and guide the user in solving the signal acquisition problems and to simplify and enhance the reliability of the diagnosis algorithms. For instance, in tele-health applications, this has a direct and significant impact on system usability, user acceptance, long-term adherence and safety. Other non-negligible consequences are on the impact of the number of interventions by professionals induced by high rates of false positive diagnosis (Cleland *et al* 2005), typically associated with tele-monitoring systems. These might interfere in ordinary clinical workflows and have non-negligible financial and motivational impacts.

Many researchers have applied electrocardiogram (ECG) as a reference or marker to find noise contaminations in HSs. In Carvalho *et al* (2005) an ECG was applied for heart beat detection. Subsequently, noise presence detection was performed by beat-to-beat power spectrum cross correlation. A very well-known method for speech enhancement based upon spectral domain minimum-mean squared error (MMSE) estimation was applied to reduce noise effects in HS (Paul *et al* 2006). This method reduces white noise from HS, while S3 and S4 sounds were prevented using ECG gating. The problem of noise detection is explicitly addressed in Barschdorff *et al* (1995) and Rajan *et al* (1998). In Barschdorff *et al* (1995), the noise detection process starts by dividing systole into two segments, containing the first and second HS, respectively. In each segment the variance is computed. All systoles with variance in either segment significantly above a threshold value are discarded as being noisy. The threshold value used depends on the smallest variance calculated across all systoles. The method described in Rajan *et al* (1998) detects noise by passing the wavelet coefficients of half-second windows of audio to single layer perceptrons that have been previously trained on clean and noisy data. Another attempted method for noise cancellation in real time was developed using an extra acoustic sensor to capture the environmental noise. This additional signal provides a noise signal for subtracting environmental noise from the contaminated HS signal (Bai and Lu 2005). Other methods can be found in the literature which involve filtering with a certain band of frequencies (Brusco and Nazeran 2005). One might look for inspiration for solving this noise detection and cancelation problem in the well-established speech recognition literature. In this area of research, it is observed that noise cancelation is typically tackled using one of two main approaches: exploring the intrinsic characteristics of vocal sound and applying blind source separation (BSS) techniques. The standard formulation of blind source separation (e.g. using ICA—independent component analysis—or the phases between different signals) requires at least as many sensors as signal sources (Comon 1994, Kristjansson *et al* 2004). Extensions exist on BSS for speaker separation, which are not bounded by this independent number of observations constraint. These solutions typically rely on constraints inspired on the characteristics of voice, such as the regular harmonic structure of voiced phonemes (Zhang and Zhang 2006), *a priori* knowledge on sound models (Kristjansson *et al* 2004, Potamitis and Ozerov 2008, Jang and Lee 2004) or by assuming that fundamental frequencies do not overlap (Barry *et al* 2005). In the context of HS processing these approaches have

some disadvantages. In order to keep the monitoring system cost effective, practical and user friendly, it is recommendable to limit the number of signals to be acquired in home settings. On the other hand, it is observed that practical results achieved by existing BSS techniques are far from perfect in terms of separation error, i.e. signal separation errors cannot be neglected and might compromise important diagnosis characteristics in HSs. These separation errors tend to deteriorate further in single source BSS algorithms. Furthermore, *a priori* knowledge typically required by single source BSS techniques is difficult to achieve for the problem at hand. For instance, HSs do not exhibit a harmonic structure. Another aspect that is not conveniently tackled by most of these noise detection approaches is their resilience with respect to the person-specific collection context (site, posture, system characteristics) and specific HS morphologies.

In most health applications involving HS collection, it suffices to detect noise occurrence and to discard contaminated sound clips or, if the noise persists, to alert the user in order to take the necessary actions to eliminate the problem. The reason for this is that, apart from rare persistent and quasi-periodic noise sources such as heavy lung sounds induced by respiratory diseases, most noise sources are transient and/or can be easily avoided. This is the strategy proposed in this paper, i.e. in section 2 an algorithm is proposed for non-cardiac sound detection during real-time signal recording that is able to detect transient or persistent non-quasi-periodic noise sources. In the presence of quasi-periodic contaminations such as respiration noise, the proposed strategy is still useful to obtain an adequate HS recording that can further be processed to eliminate the respiratory sounds, such as the one proposed by Tsalaile (2008).

This algorithm has been tested using a database of 115 HS clips that contain representative noise contaminations that have been acquired from 71 patients with several types of coronary and heart diseases. In order to thoroughly evaluate the algorithm's performance on several types of additive noise sources, a simulation study is also introduced. These results and their discussion are presented in section 3. The main conclusions of the proposed noise detection approach are introduced in section 4.

2. Method

Figure 1 depicts the proposed approach for the detection of HS segments with noise contamination. As can be observed, the method is composed by two main phases, denoted by phase I and phase II, respectively. The goal of phase I is to find a segment of uncontaminated heart HS correspondent to one complete heart cycle. This heart cycle will be treated as a reference sound in phase II of the algorithm.

In phase I, given a window of HS, first the algorithm extracts the heart rate from the envelope of the HS. This is performed resorting to the autocorrelation function and corresponds to step (A1) described next. The estimated heart cycle duration is then applied to segment individual heart cycles, which are then assessed for noise contaminations in the time domain (see left flow of phase I in figure 1). The later is based on the observation that HS exhibits quasi-periodic behaviour in the time domain, which is assessed using a similarity measure between adjacent heart beats (see step A2). This quasi-periodicity behaviour of HS also manifests itself in the time-frequency domain. Namely, it is observed that the time distribution of the energy in adjacent frequency bands tends to be (i) highly correlated and (ii) their peaks tend to be aligned. These conditions are inspected in the processing steps on the right path of the flow chart of phase I in figure 1 and will be described in steps (B1) and (B2), respectively. The heart cycles, which meet simultaneously the temporal similarity as well as the spectral
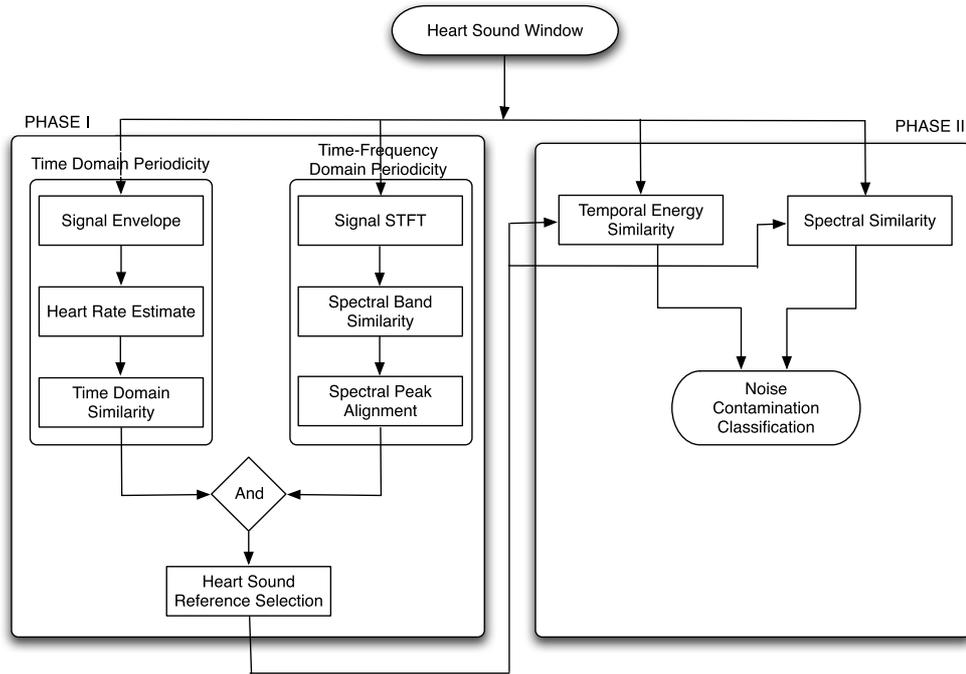
Figure 1. Flow chart of the proposed approach.

similarity criteria are taken as candidates for the reference HS template. The actual template is selected among this set of candidates.

In phase II (see the right path of the flow chart in figure 1), each acquired HS window is checked for similarity against the selected HS reference template. Here two criteria have to be fulfilled simultaneously: (i) spectral similarity and (ii) temporal energy similarity. The later is taken into account in order to determine short duration sound spikes, which are not captured by the spectral similarity criterion.

2.1. Phase I: Reference sound detection

2.1.1. Detection of template candidates. This section introduces the algorithm steps required to identify candidates for the HS template.

(A) Periodicity in the time domain. The envelopes of HS components are extracted by applying the Hilbert transform followed by the Gammatone band-pass filter (Deshmukh *et al* 2005). Next, the autocorrelation function of the envelope is computed. Typically, it will exhibit pronounced peaks for the main HS components, i.e. the S1, S2 and murmur components. The autocorrelated values are normalized by the autocorrelation values of a chosen windowing function, e.g. the Hanning window. Let $x^e(t)$ be the envelope of the HS, $x(t)$, and let $y(\tau)$ be its autocorrelated function such that

$$y(\tau) = \frac{\int_{-\infty}^{\infty} x^e(t)w(t)x^e(t-\tau)w(t-\tau)\,dt}{\int_{-\infty}^{\infty} w(t)w(t-\tau)\,dt} \tag{1}$$

where $w(t)$ is the Hanning window function and T is the duration of a given segment of HS, which is of the length of 4 s. The variable τ denotes lag in computation of the autocorrelation that varies from zero to the duration of the analysed signal window. Recently, Schmidt *et al* (2010) and Jimne-Gonzlez and James (2010) have used autocorrelation for periodicity identification in HS segmentation and in the extraction of periodic signals from noisy abdominal phonograms, respectively. Herein, we also propose to use the autocorrelation function to identify the heart cycles present in the analysis window. However, this cannot be done directly by identifying the prominent peaks in $y(\tau)$, since the prominence of the autocorrelation peaks tend to be significantly disturbed in the presence of noise and murmur. Even in clear HSs, it is observed that usually false prominent peaks are detected. To solve this, the number of prominent peaks in $y(\tau)$ are constrained using the estimated heart rate. The actual algorithm to select the prominent peaks, which define the heart cycles, is based on an adaptation of Park's method (Park 2000), i.e. (i) first the peak with the highest amplitude is selected; (ii) the algorithm searches the adjacent prominent peak to the left using a window whose duration equals the heart cycle period; the peak with the highest amplitude inside that window is selected; (iii) the same procedure described in step (ii) is repeated with a window to the right; (iv) steps (ii) and (iii) are recursively repeated using the selected prominent peaks in the previous iteration; (v) the algorithm stops when the number of selected prominent peaks equals the number of heart cycles inside the analysis window (these are calculated using the estimated heart rate and the duration of the window). Once all prominent peaks have been detected, the periodicity is checked using the radial distance (4) between two contiguous heart cycles. The heart rate estimation and periodic cycles verification procedures are as follows.

(A1) Heart rate estimation. In the absence of an auxiliary signal, such as the ECG, it is observed that the heart rate can be estimated using the first two singular values of the rearranged autocorrelation function of the envelopes of the segment of HS under analysis by extending the algorithm described in Kanjilal and Palit (1994). Let $y(\tau)$ be periodic with the period T', i.e. $y(\tau + T') = y(\tau)$. For the typical population at rest, it is observed that T' is between 500 and 1200 ms (heart rate is between 120 and 50 beats min^{-1}). Let $Y(T')$ be the data matrix which is constructed by stacking $y(\tau)$ samples after every T' ms using the following arrangement:

$$Y(T') = \begin{pmatrix} y(0) \ldots y(T') \\ y(T' + T_s) \ldots y(2T') \\ \cdot \\ \cdot \\ y((m-1)T' + T_s) \ldots y(mT') \end{pmatrix} \tag{2}$$

where m is the number of periodic analysis segments in a given segment of HS, and T_s is the sampling period i.e. time interval between two consecutive lags in $y(\tau)$. If $Y(T')$ has linear dependent rows, then it will have some zero singular values. The singular values are found by singular value decomposition (SVD) of $Y(T')$. Let $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_3 \ldots \geqslant \sigma_m$ be the aforementioned singular values. The periodicity can be measured using (3):

$$\rho = (\sigma_2/\sigma_1)^2. \tag{3}$$

A value of ρ near zero implies a strong periodicity in the signal. In the process of estimating the heart rate, $Y(T')$ is constructed by varying T'. The T' which minimizes ρ, is the estimated duration of the heart cycle. The estimated heart rate enables us to find the prominent peaks in $y(\tau)$. Since prominent peaks are directly related to the main components in the HS, which occur only once per heart cycle (see figure 2), this is used as the prior information to find peaks.
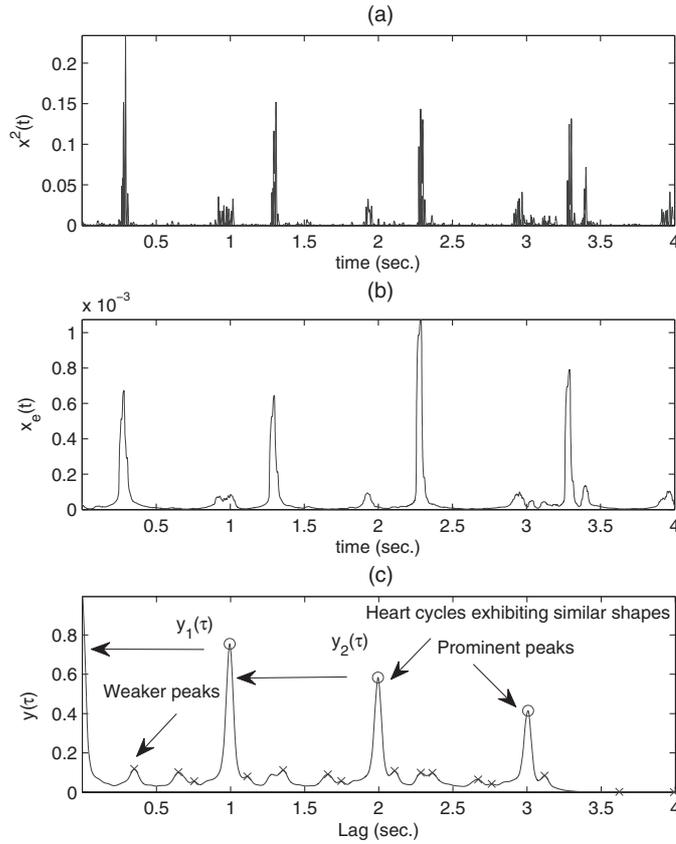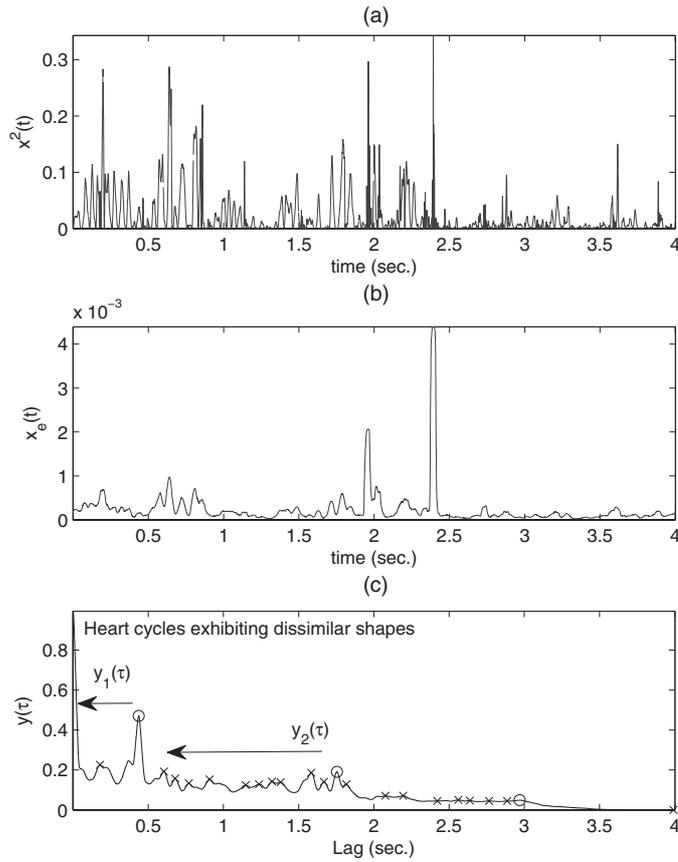
Figure 2. (a) Heart sound energy, (b) heart sound envelope, and (c) autocorrelation function ($y(\tau)$) of the envelope with peak identification (showing similar heart cycles in terms of vectors $y_1(t)$ and $y_2(t)$).

Namely, inside each processing window, first the largest peak in amplitude is identified. Then the algorithm determines the second largest peak at a distance not greater than the identified heart cycle duration from the first peak. This approach is repeated for each newly identified peak until the total number of prominent peaks found equals the number of complete heart cycles in the processing window.

(A2) Periodicity check criterion. The set of prominent peaks in $y(\tau)$, which has been detected in the previous step, enables us to find shape similarity between two heart cycles (contiguous pair of prominent peaks). Since the prominent peak detection algorithm is constrained by the estimated heart rate, it is observed that each pair of adjacent prominent peaks defines a complete heart cycle. The similarity is measured by the radial distance between two vectors. Let $y_r(\tau)$ and $y_{r+1}(\tau)$ be the vectors representing the portions of the autocorrelation function of two consecutive heart beats; then the radial distance is given by

$$\mathrm{Cos}(\theta) = \frac{\langle y_r(\tau), y_{r+1}(\tau)\rangle}{|y_r(\tau)||y_{r+1}(\tau)|} \tag{4}$$

Figure 3. (a) Noisy heart sound, (b) heart sound envelope, and (c) autocorrelation function $y(\tau)$ (showing dissimilar heart cycles).

where $\langle \cdot \rangle$ is the inner product operator and $| \cdot |$ represents Euclidean norm of the vectors. In (4), $y_r(\tau)$ and $y_{r+1}(\tau)$, where r is the number of prominent peaks in the analysis window, are interpolated if they do not have the same length. It is considered that two consecutive heart cycles exhibit similar shapes if their internal product is greater than 0.8. This threshold was tuned experimentally. Similar shaped and dissimilar shaped heart cycles in clean and noisy HSs can be observed in figures 2(c) and 3(c), respectively.

(B) Periodicity in the time-frequency domain. The periodicity check of HSs in the time domain is not sensitive to the presence of many non-cardiac sounds. For instance, swallowing, breathing or high pitched voice cannot be identified using the time domain periodicity detection technique. However, the influence of the noise source in HS periodicity can be observed in the time-frequency domain. In the proposed method, the spectrogram is adopted to find periodic patterns in the time-frequency bands. Let $S(f, t)$ be the short-time Fourier transform (STFT) of $x(t)$, i.e.

$$S(f, t) = \int_{-\infty}^{\infty} x(\eta) w(\eta - t) \exp(-2jf\pi\eta) \, d\eta \qquad (5)$$
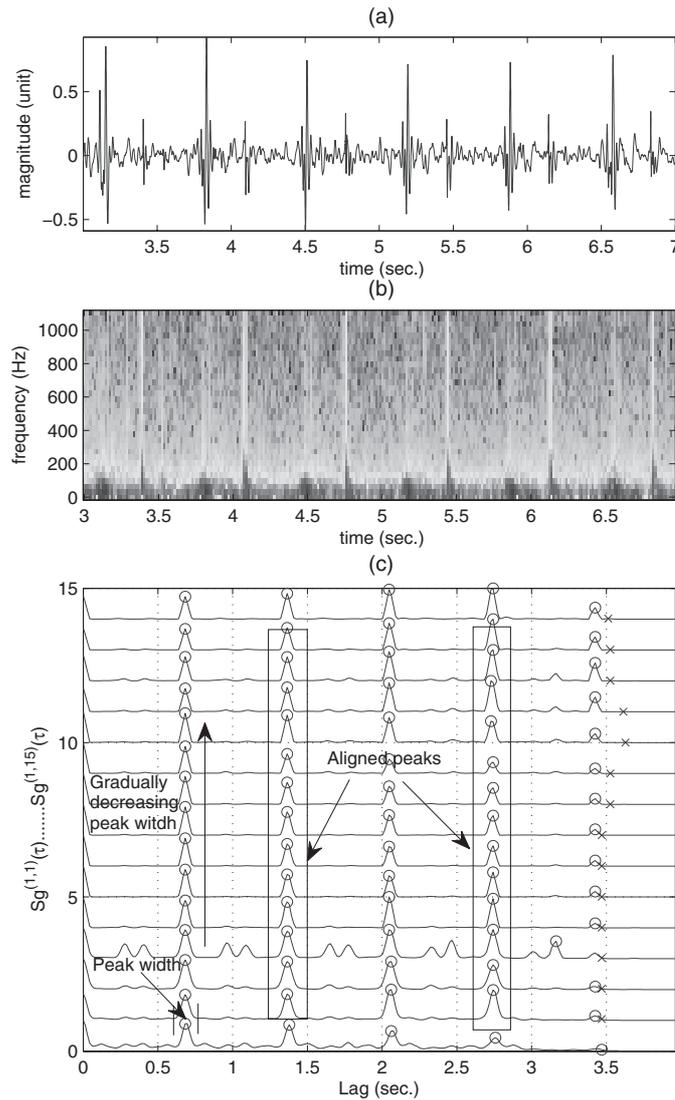
Figure 4. (a) Normal heart sound from a single-tilted disc mechanical prosthetic valve in the mitral position, (b) spectrogram of (a), and (c) the autocorrelation function $AS^k(\tau)$ of spectral power in 15 frequency bands.

where $w(\eta)$ is the Hanning window function. The windowing function satisfies $w(\eta) = 0$ for $\eta > T/2$, where T, as mentioned in (A), is the length of the analysed HS segment. In HSs produced by healthy subjects as well as by individuals who exhibit cardiac diseases, it is observed that most of the energy of the HS signal is concentrated in the 0–600 Hz frequency range. In the proposed approach, this frequency range is divided into 15 evenly divided frequency bands and the spectral energy in each of the bands is taken for periodicity validation. In figure 4(c) the autocorrelation functions of the energy in each of these 15 frequency bands are depicted for a normal heart sound. As can be observed, the heart sounds exhibit regular

patterns in the autocorrelation functions computed for each frequency band. Furthermore, these regular patterns tend to be linearly dependent between increasing frequency bands. These linear dependences may monotonically increase or decrease depending on the type of heart sound. Namely, it is seen that it tends to decrease for heart sounds associated with natural or bioprosthetic heart valves, while for heart sounds produced by mechanical prosthetic heart valves it is observed that the linear dependency tends to increase with increasing frequency. The later is associated with the fact that heart sounds induced by mechanical heart valves exhibit much higher frequency content compared to natural or bioprosthetic heart valves. Furthermore, it is also seen that the peaks of the autocorrelation functions in each frequency band tend to be aligned in time. This is due to the fact that most of the signal's power is due to the S1 and the S2 heart sound components which are responsible for the main peaks in the autocorrelation functions.

The methods for the verification of S1 and S2 periodicity in the time frequency domain using the 15 frequency bands and the criterion for peaks alignment in these bands are explained next.

(B1) Pattern detection in frequency bands. In order to extract the periodic patterns from the spectrogram, autocorrelation of the energy of each frequency band is obtained from (1). Let $AS^k(\tau)$ be the autocorrelated function of the kth, where $k = 1, 2, 3 \ldots 15$, frequency band in the spectrogram. As can be observed in figure 4(c) the autocorrelated power in 1–15 frequency bands are in regular patterns, where the prominent peaks occur almost aligned in time. Furthermore, it is seen that the widths of these peaks tend to decrease in higher frequency bands (see in figure 4(c)). These observations are the basis to build the heuristic which is applied to verify if the cardiac signal is clean from noise contamination. One of the observations is the linear dependence of the rows of $AS^k(\tau)$. To evaluate linear dependency contiguous ascending frequency bands are grouped according to (6) and singular values are computed using the previously described SVD technique:

$$Sg^{(m,m+4)}(\tau) = \begin{pmatrix} AS^m(\tau) \\ AS^{m+1}(\tau) \\ \cdot \\ \cdot \\ AS^{m+4}(\tau) \end{pmatrix}, \qquad m = 1, 6, \text{ and } 11. \tag{6}$$

In equation (6) $Sg^{(m,m+4)}$ is the matrix formed by grouping the $AS^k(\tau)$ rows for each five contiguous frequency bands. The linear dependence is assessed using (3), since ρ should be near zero for linear dependent rows. Regarding the monotony assessment of the linear dependence in increasing or decreasing frequency bands, ρ may also assist in its verification. Let ρ_1, ρ_2 and ρ_3 be the singular values ratios computed using (3) with the eigenvalues calculated from matrixes $Sg^{(1,5)}(\tau)$, $Sg^{(6,10)}(\tau)$ and $Sg^{(11,15)}(\tau)$, respectively, then the most significant observations regarding pure HSs are: $\rho_1 > \rho_2 > \rho_3$ or $\rho_1 < \rho_2 < \rho_3$.

In HSs, two distinct situations are observed for (i) native and bio-prosthetic valves and (ii) for mechanical prosthetic valves. For native valves or bio-prosthetic valves, it is seen that the main signal energy is concentrated in the lower frequency spectrum. Higher frequency components appear only at very short time periods (e.g. at the onsets of the aortic valve closing in S2). Furthermore, the spectral range of HS produced by natural and bioprosthetic valves is lower compared to the spectral range of HS produced by mechanical prosthetic valves. Hence, the autocorrelation function is less regular as frequency increases (see figure 5). This leads to the lower linear dependency of the autocorrelation function as frequency bands increase, i.e. $\rho_1 < \rho_2 < \rho_3$. Another situation is found in mechanical prosthetic valves. In this situation,
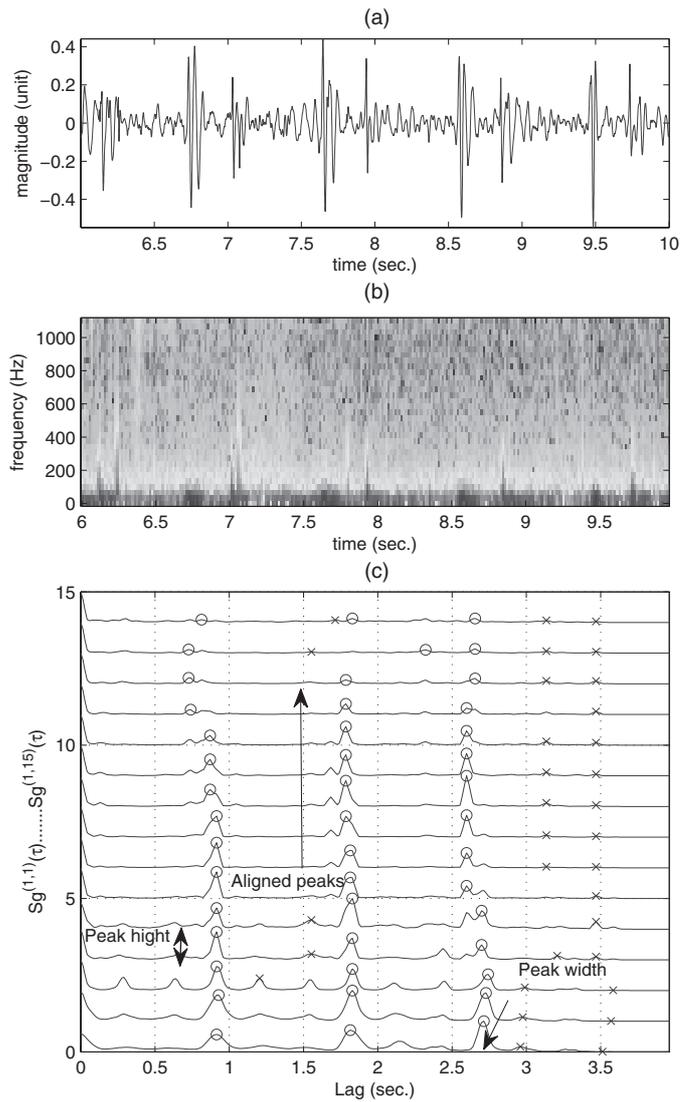
Figure 5. (a) Normal heart sound from a native valve, (b) spectrogram of (a), and (c) the autocorrelation function $AS^k(\tau)$ of spectral power in 15 frequency bands.

it is observed that the higher frequency bands have considerably more energy in a broader frequency range, leading to more pronounced and better defined peaks in time (i.e. peaks of considerable amplitude and very well localized in time) in high frequency bands. Hence, the linear dependency increases in this situation, i.e. $\rho_1 > \rho_2 > \rho_3$. On the contrary, these two conditions are not met in the presence of noise (see an example in figure 6).

(B2) Peak alignment in the frequency bands. As has already been explained, the main autocorrelation peaks are due to S1 and S2 HSs. Therefore, they should exhibit alignment in
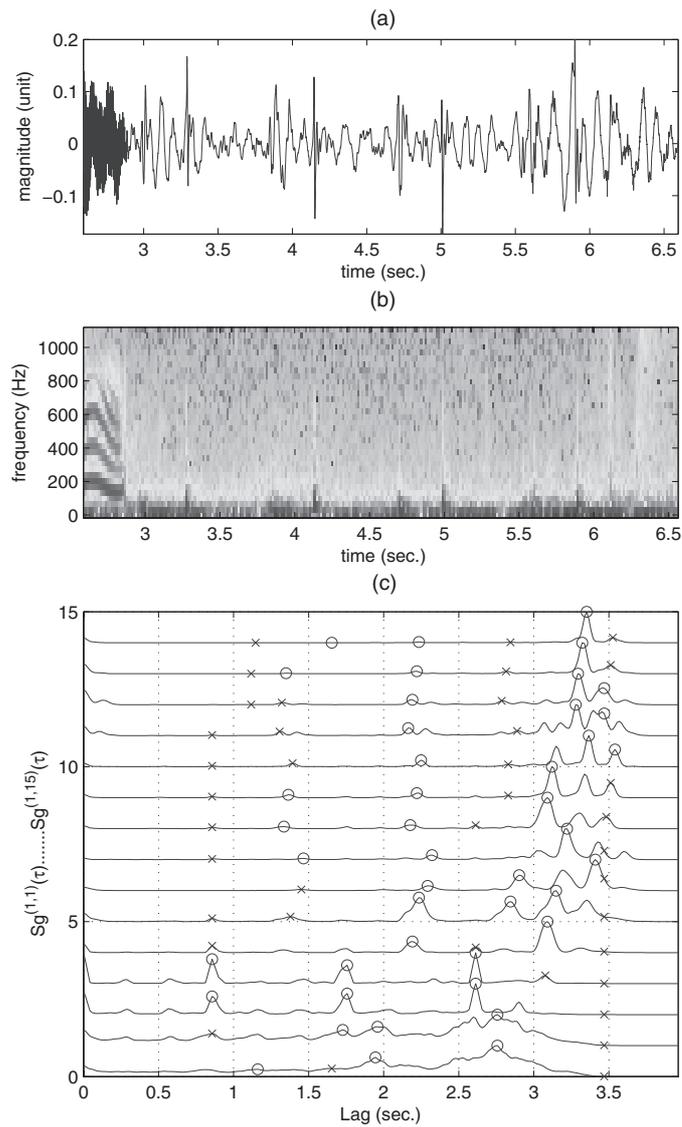
Figure 6. (a) Noisy heart sound from a native valve, (b) spectrogram of (a), and (c) autocorrelation function $AS^k(\tau)$ of spectral power in 15 frequency bands (exhibiting aperiodicity).

$Sg^{(m,m+4)}(\tau)$ frequency bands. Otherwise, it is considered that these peaks are due to noise. In order to check the alignment, all main peaks are found using the previously described peak detection technique. Afterwards, defining a time tolerance ($\pm 5\%$ of the time of the peak in the first frequency band), the alignment of all peaks is inspected (see an example of aligned peaks in figure 4(c)). Although these peaks do not follow regular alignment in the presence of noise, as can be see in figure 6(c). Finally, if 80% of the total peaks are detected aligned and the monotonicity criterion is met, then the analysis window is assessed as an uncontaminated HS segment.

2.1.2. A heart cycle selection as the reference sound. The processing steps described in the previous subsection enable the identification of uncontaminated HSs. These steps are applied to identify a window containing clear HSs that will serve to extract a HS reference template. In order to achieve this goal, first complete heart cycles are identified inside the window using the autocorrelation function peaks. Second, the heart cycle with the highest average similarity (radial distance) with respect to all available heart cycles in the window is selected as the reference heart cycle template.

In the current implementation of the algorithm, a window of 4 s of HS is taken to estimate the reference HS. This analysis window was found to be sufficient to examine the signal's periodicity in the time and in the time-frequency domain. If the signal satisfies the periodicity conditions introduced in subsection 2.1, a reference HS is extracted using the aforementioned procedure. Otherwise, the window is shifted 1 s forward and the process is repeated until the periodicity conditions are met.

2.2. Phase II: non-cardiac sound detection

The previous subsections introduced the preparation phase which is required to detect the reference HS. Once the reference HS has been defined, a template-matching approach is applied using the following spectral and temporal features.

2.2.1. Spectral energy. In this step, the spectral root mean square of the HS signal is calculated, i.e.

$$S_{\mathrm{rms}}(f) = \sqrt{\int_t^{t+T_h} |S(f,t)|^2 \, \mathrm{d}t} \tag{7}$$

where T_h is the length of the window of the HS which has the same duration as the reference sound (duration of the estimated heart cycle). Root mean square of the spectrogram provides an estimate of the power distribution in the frequency domain. Let $S_{\mathrm{rms}}^{\mathrm{ref}}(f)$ and $S_{\mathrm{rms}}^{\mathrm{test}}(f)$ be the spectral root mean square for the reference and the test HS signals respectively; then validation is performed using the following condition:

$$\mathrm{CorrCoef}\big(S_{\mathrm{rms}}^{\mathrm{ref}}(f), S_{\mathrm{rms}}^{\mathrm{test}}(f)\big) > th_1 \tag{8}$$

where CorrCoef is the correlation coefficient between two signals, and threshold th_1 value is defined as 0.98. A HS segment is assessed as noisy if CorrCoef goes below th_1 in (8). An example of the effects of (8) in assessing HS contamination by several sources is depicted in figure 7(b).

2.2.2. Temporal energy. Temporal energy is taken into account in order to determine short duration sound spikes. It is seen that spectral energy correlation is incapable to capture these types of noises with the defined correlation criterion in (8), see in figure 7(c). However, it can be efficiently tackled with the temporal energy of the sounds.

In order to obtain a significant information about the instantaneous amplitude of non-cardiac sounds of a short duration, it is important to compute the energy for small windows of the test signal. In our experiments we use window of duration $t_w = 50$ ms. The instantaneous energy is compared with the maximum energy of the reference sound computed in successive t_w window according to (9). Let $x^{\mathrm{ref}}(t)$ be the reference signal and $x^{\mathrm{test}}(t)$ be the sound under
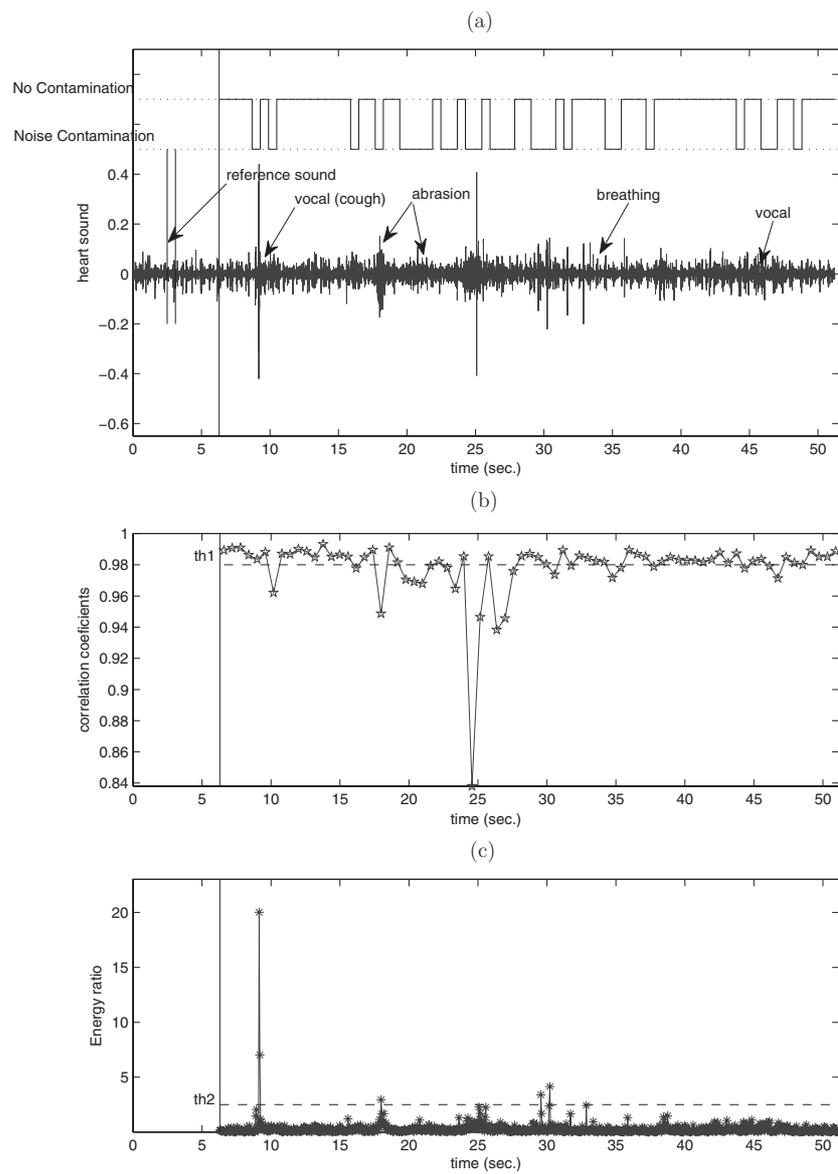
Figure 7. (a) Heart sound contaminated with several sources of noises. First 6 s correspond to reference template search window. The upper part of the graph shows noise classification result. (b) Spectral energy correlation between heart sound segments and the reference sound. (c) Relative temporal energy of the heart sound segments.

Table 1. Induced noise sources.

Category of noises	Noise type
Vocal	Speech at different pitches
	Coughing
	Laughing
Physiological	Breathing /heavy breathing
	Swallowing
	Muscle movement
Sensor	Rubbing
	Movement
Ambient	Knocking on the door
	Noises induced by moving/falling of small objects
	Ambient music
	Electric table fan
	Phone ringing
	Foot steps

assessment. The relative temporal energy (RTE) of the lth segment under analysis, is defined in (10):

$$TE^{ref} = \max_{l=1,2...,n_p} \int_{(l-1)t_w}^{lt_w} x^{ref}(t)^2 \, dt \tag{9}$$

$$RTE_l = \frac{\int_{(l-1)t_w}^{lt_w} x^{test}(t)^2 \, dt}{TE^{ref}} \tag{10}$$

$$x^{test}(t) = \begin{cases} \text{noise} & \exists l \in \{1, 2, \ldots, n_p\} : RTE_l > th_2 \\ \text{heart sound} & \text{otherwise} \end{cases} . \tag{11}$$

In (10) t_w is the window duration applied to compute the signal's energy, while $n_p = [\frac{T_h}{t_w}]$, $l = 1, 2, 3, \ldots, n_p$, is the total number of sound segments in the test sound. A sound segment of duration T_h is identified as contaminated with spike or impulse sounds of short duration, as formulated in (11), if its corresponding RTE exceeds the threshold th_2, where the threshold is set to 3-fold TE^{ref}.

2.3. Material and data collection

Heart sounds were recorded from patients with prosthetic valve implants (both mechanical and bioprosthetic) one month after valve implant surgery as well as some from healthy volunteers. A commercial stethoscope from Meditron has been utilized for HS acquisition. The device has an excellent signal-to-noise ratio and an extended frequency range (20–20 000 Hz). All sound samples were digitized with 16-bit resolution and 44.1 kHz sampling rate.

The prepared data set includes 115 HS clips of recording length between 1 and 2 min. These HS clips have been collected from 71 different subject at rest with the following biometric characteristics (average±standard deviation): age = 35.26 ± 12.02 years; BMI = 25.11 ± 7.8 kg m^{-2}; male subjects = 64; female subjects = 7. Regarding heart dysfunction, three subjects exhibited arrhythmia, 31 subjects had an artificial valve implant, andeight patients exhibited heart murmurs and the rest were healthy subjects.

In order to validate the performance of the algorithm, several types of noises were intentionally induced during the HS acquisition protocol. According to table 1, in the prepared

database, non-cardiac sounds were arranged into four classes: vocal sounds, physiological sounds (breathing, coughing, etc), skin rubbing or abrasion sounds via the stethoscope (sensor based), and other ambient noises. Contaminated HSs have been applied during the tests of both phases of the algorithm. The collection of HSs applied for the validation of the second phase of the method was composed by 4781 uncontaminated HS beats and 1659 contaminated heart beats distributed as follows: 805 vocal, 241 sensor, 161 physiological and 452 ambient noise segments. It should be mentioned that noise segments were manually annotated under the supervision of a clinical expert. Furthermore, noise contaminations induced accidently during the acquisition protocol were also annotated and included in the test database.

2.4. Performance assessment

The performance of the algorithm is computed in the form of sensitivity (SE) and specificity (SP) measures, i.e.

$$\text{SE}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \; ; \quad \text{SP}(\%) = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12}$$

TP denotes the number of noisy segments correctly detected as noisy segment, TN represents clean HS segments correctly detected as clean HS segments, FP stands for clean HS segments incorrectly detected as noisy segment, and FN denotes the number of noisy segments detected as clean HS segment.

In order to evaluate the robustness of the proposed algorithm, two types of tests have been performed: in the first test the algorithm has been applied directly to the test database; in the second test set a simulation study has been performed in order to evaluate its sensitivity with respect to noise the intensity level. In the first test, noisy segments are divided based upon their loudness into low, medium and high intensity noise classes. The loudness is defined herein as the average power of the signal in decibel, i.e. let $x(t)$ be the HS mixed with noise of duration T; then the loudness can be described in (13):

$$\text{loudness (dB)} = 20 \log_{10} \sqrt{\frac{\int_0^T x^2(t)\,dt}{T}}. \tag{13}$$

In the second test set performed, uncontaminated HSs have been artificially mixed additively with several types of noises with modulated intensities. Namely, let $v(t)$ be the noisy sound in the simulation study the algorithm has been applied to $u(t) = x(t) + Kv(t)$, $K = 0, 0.5, 1, 1.5, \ldots, 4$. For $v(t)$ noise sources of the same categories as the ones described in section 2.3 have been applied. Furthermore, the signal-to-noise ratio is computed according to (14):

$$\text{SNR (dB)} = 20 \log_{10} \frac{\int_t^{t+T} x(t)^2\,dt}{\int_t^{t+T} K^2 v(t)^2\,dt} \tag{14}$$

2.5. Experimental results

The first phase of the algorithm is related to the estimation of the reference HS. In the applied database, the proposed methodology was always able to detect a reference HS. On average, the reference HS was adequately detected during the first 11.4 s (between 4 and 25 s) of each sound clip. Regarding the processing time complexity of the algorithm, it is observed that the first phase of the algorithm is significantly more time consuming than the second one. The later one might be performed in real time in most processing platforms. In test performed using an implementation in Matlab 7.6 running on Windows XP using a Intel® Core™2Duo

Table 2. Results of noise detection in terms of sensitivity (*SE* in %).

Noise type	Heart sound/ Valve type	Sensitivity			Noise sensitivity	Average sensitivity
		HL	ML	LL		
Vocal	Native valve	98.34	94.99	89.04	96.78	96.62
	Prosthetic valve	95.45	94.32	100	94.64	
	Murmur	97.06	96	100	96.74	
	Arrythmia	100	100	100	100	
Sensor	Native valve	100	97.95	92	97.16	96.43
	Prosthetic valve	100	90	90.91	92.68	
	Murmur	100	100	50.00	90.91	
	Arrythmia	100	100	100	100	
Physiological	Native valve	100	89.66	92.08	92.20	92.94
	Prosthetic valve	100	100	91.67	92.87	
	Murmur	100	100	100	100	
	Arrythmia	NA*	NA*	NA*	NA*	
Ambient	Native valve	98.37	92.95	94.45	95.24	94.47
	Prosthetic valve	100	93.33	90.00	93.18	
	Murmur	100	100	100	100	
	Arrythmia	100	80.00	66.67	77.78	
Overall sensitivity						95.88

*NA = Not available.

Table 3. Results of noise detection in terms of specificity (SP in %), computed based upon false positive (FP).

Heart sound	Specificity
Native valve	97.24
Prosthetic valve	97.61
Murmur	98.17
Arrythmia	99.47
Overall specificity	97.56

processor at 2.53 MHz, it was observed that the first phase took on average 1.23 s per each 4 s window of sound, while in the second phase the algorithm only took on average 0.035 s per processing window (duration of one heart cycle).

In order to examine the noise detection performance of the algorithm, noisy sound segments have been divided into three classes based on their degree of loudness, i.e. low loudness (LL), medium loudness (ML) and high loudness (HL), and into four classes of noise types based on the origin of the noise source. The considered ranges of loudness in each class were: HL > -7.5 dB, -14 dB $\leqslant$ ML $\leqslant -7.5$, LL < -14 dB. Tables 2 and 3 summarize the achieved detection performance results in terms of sensitivity and specificity in each of these classes of noises for distinct types of cardiac dysfunctions. As can be observed in tables 2 and 3, the overall achieved detection sensitivity and specificity are 95.88% and 97.56%, respectively. The achieved detection performance for each of the considered classes of noise contaminations is relatively homogeneous. Namely, the obtained results show that the proposed method enables the detection of vocal noises at different pitches with a sensitivity of 96.62%, irrespective of cardiac dysfunction. Regarding noise contaminations originated

by sensor movements, i.e. noise contaminations of class sensor induced noise, the method enables its detection with 96.43% sensitivity. Another highly probable incident during HS acquisition is contamination by physiological noises. As can be observed, these types of noises are detected with 92.94% sensitivity. This lower value in sensitivity is directly related to the fact that physiological noises tend to exhibit similar spectral content as regular heart sounds and, hence, are more difficult to capture using the spectral similarity feature. Regarding ambient sound contaminations, the algorithm exhibits 94.47% sensitivity. It should be noted that these classes of noises reflect the most likely noise contaminations during HS acquisition. Regarding the performance with respect to the cardiac dysfunction, table 2 suggests that the algorithm exhibits comparable performance irrespective of heart dysfunction and valve type (implant or native).

It should be noted that the results reported for arrhythmia are not significant, since only three subjects in the database exhibited arrhythmia. As one would expect, the noise detection performance deteriorates as the noise loudness decreases. Using the test database, it is seen that the algorithm exhibits a deterioration less than 10% in sensitivity when noise intensity decreases from HL to LL, irrespective of the noise type and cardiac dysfunction. As can be observed in some of the reported results that sometimes the algorithm seems to perform better for low loudness noise perturbations than for medium loudness noises. This is mainly due to the way the noise intensity is assessed (13). The average loudness is not able to adequately capture the instantaneous intensity level of short duration noises and, therefore, it is observed that some reasonably loud short duration noises might be classified in the LL class, leading to the aforementioned effects.

Table 3 summarizes the specificity results achieved by the algorithm for distinct heart dysfunctions. As can be observed, the obtained specificities tend to exhibit resilience with respect to the characteristics of the HS source. Given the high values of SP experienced under the method, a very low number of false positives are expected, i.e. the method tends to exhibit a very small number of regular HS segments that are classified as noisy segments.

In order to analyse the noise intensity implications in the algorithm's performance, a simulation study was setup using an additive noise contamination model with amplitude modulation. In this study, clean HSs acquired from subjects with native heart valves, prosthetic heart valves, murmur and arrhythmia were additively mixed with noise contaminations of the considered noise classes, i.e. voice, sensor, physiological and ambient noises. During this process, the SNR of the HS segments mixed with noise sources was varied using a linear gain. The achieved results are reported in figure 8. As can be observed, regardless of the type of heart dysfunction and noise contamination, the sensitivity of the algorithm declines as the SNR increases. For negative SNR values, i.e. situations where the noise intensity exceeds the signal energy, the sensitivity of the method is relatively stable and very high with SE values near 100%. At 0 dB SNR, a situation where the noise exhibits the same energy as the signal, it is observed that the performance of the algorithm exhibits a SE over 85%–90%. As one would expect, since the energy of the noise decreases below the energy of the signal (positive SNR region) the number of false negatives increases rapidly, leading to a drop in sensitivity. However, it should be mentioned that even for very low energy noise contaminations, the algorithm exhibits a significant behaviour. For instance, for 20 dB noise contamination, i.e. situations where the noise intensity is 10% of the energy of the signal, the simulation results suggest that the method exhibits SE values higher than 80% for most noise contamination and HS types. Regarding the performance of the algorithm with respect to the contamination type, it should be observed that the detection sensitivity tends to be significantly more consistent and higher for larger ranges of SNR in noise contaminations of ambient noises and vocal noises. These types of noises tend to exhibit significant content in higher frequencies (e.g. vocal sounds
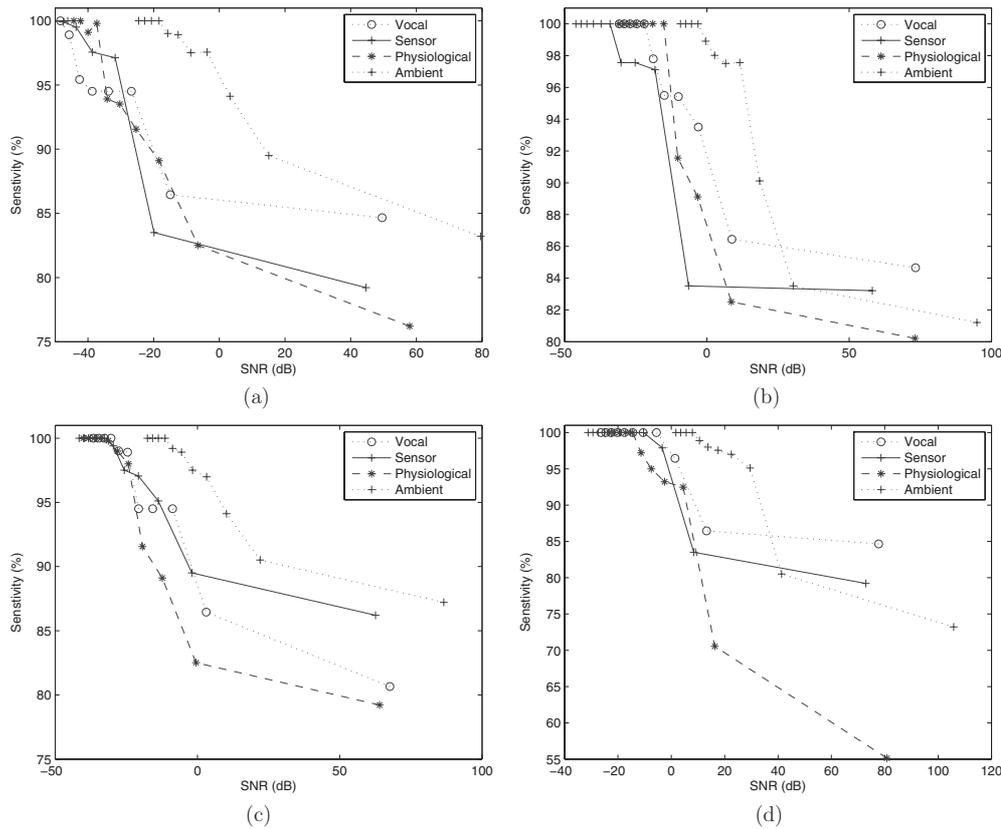
Figure 8. Sensitivity for noise detection in varying intensities of noise of the (a) heart sound from native valves, (b) prosthetic valve, (c) murmur, and (d) arrythmia.

are typically around 2–4 kHz bandwidth) and, therefore, exhibit spectral signatures that are easier to be captured by the spectral similarity test. Using the additive simulation study, it is seen that the performance of the algorithm tends to drop more rapidly for physiological and sensor noise contaminations compared to the other groups of noises. This is related to different reasons; physiological noise sources tend to exhibit similar spectral content range compared to HSs; hence, their spectral power signatures are closer to HS spectral power signatures. As for sensor noise sources, it is observed that these are usually characterized by short durations, leading to less well-defined spectral signatures. In this type of contaminations the temporal energy feature plays a central role.

3. Conclusions

Heart sound is a very informative biosignal, since it directly encodes the mechanical activity of the heart. It has been shown that HS has the potential to be applied in very important biomedical applications both in hospital as well as in home-care setting. The collection of an uncontaminated and adequate signal for analysis is probably one of the first challenges that has to be solved in designing biomedical systems based on this signal. In this paper, an algorithm has been proposed that enables the detection of HS contaminations by several sources, irrespective of their spectral and intensity characteristics and interference model.

The method is based on a template-based strategy composed by two main phases: in a first phase, the algorithm detects a reference HS with the duration of one heart cycle. This enables resilience with respect to site of collection, subject and sensor positioning as well as subject specific morphological characteristics and sound transmission paths. In a second phase, the detected reference HS template is applied in real time to detect deviations in the HS pattern. The first phase is based on two periodicity characteristics of uncontaminated HS: HS is a quasi-stationary signal exhibiting periodicity in time and it is also observed that most of its energy is concentrated in its S1 and S2 components which manifests itself in adjacent frequency bands repetitions of the same energy pattern, leading to a periodicity pattern in frequency bands. These two observations are applied to identify an uncontaminated HS heart cycle. It should be mentioned that this processing step is the most demanding one from the computational load perspective. Once this template has been identified, the algorithm proceeds in detecting non-cardiac sounds by checking the template against the acquired sound. Herein, two computationally simple features are extracted to analyse the similarity with respect to the template. Simulation tests using Matlab suggest that this phase of processing can be performed in real time enabling the algorithm to be used to provide feedback information to the user regarding the quality of the signal.

The method was exhaustively tested using a database of HSs collected from 71 subjects with different cardiac dysfunctions and noise contaminations. The database has a total of 157 min of collected HSs with clean HSs as well as HSs that have been contaminated with some of the most probable noise sources during real life HS acquisitions at several intensity levels. Using this database, the algorithm achieved a sensitivity of 95.88% and a specificity of 97.56%. These high sensitivity and specificity together with its real time operation, makes this algorithm an interesting solution to deploy HS-based biomedical systems, particularly in pHealth scenarios where feedback information has to be provided in real time to the user in order to allow us to timely alert and guide the user in solving the signal acquisition problems and to avoid false positives.

Acknowledgments

References

Abbas A and Bassam R 2009 *Phonocardiography Signal Processing (Synthesis Lectures on Biomedical Engineering)* (San Rafael, CA: Morgan and Claypool Publishers)

Bai Y W and Lu C L 2005 The embedded digital stethoscope uses the adaptive noise cancellation filter and the type I Chebyshev IIR bandpass filter to reduce the noise of the heart sound *Proc. Int. Workshop on Enterprise Networking and Computing in Healthcare Industry, HEALTHCOM* pp 278–81

Barry D, Fitzgerald D, Coyle E and Lawlor B 2005 Single channel source separation using short-term independent component analysis *119th Audio Engineering Society Convention (New York, 7–10 October 2005)*

Barschdorff D, Femmer U and Trowitzsch E 1995 Automatic phonocardiogram signal analysis in infants based on wavelet transforms and artificial neural networks *Proc. Computers in Cardiology* pp 753–6

Brusco M and Nazeran H 2005 Development of an intelligent PDA-based wearable digital phonocardogram *Proc. 27th IEEE EMBS Annu. Int. Conf.* pp 3506–8

Carvalho P, Gil P, Henriques J, Antunes M and Eugénio L 2005 Low complexity algorithm for heart sound segmentation using the variance fractal dimension *Proc. IEEE Int. Symp. on Intelligent Signal Processing* pp 593–5

Cleland J G, Louis A A, Rigby A S, Janssens U and Balk A H 2005 Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the trans-European network-home-care management system (ten-hms) study *Am. Coll. Cardiol.* **45** 1654–64

Comon P 1994 Independent component analysis. a new concept? *Signal Process.* **36** 287–314

Deshmukh O, Espy-Wilson C Y, Salomon A and Singh J 2005 Use of temporal information: detection of periodicity, aperiodicity, and pitch in speech *IEEE Trans. Speech Audio Process.* **13** 776–86

Durand L G and Pibarot P 1995 Digital signal processing of the phonocardiogram: review of the most recent advancements *Crit. Rev. Biomed. Eng.* **23** 163–219

Jang G and Lee T 2004 A maximum likelihood approach to single-channel source separation *J. Mach. Learn. Res.* **4** 1365–92

Jimne-Gonzlez A and James C J 2010 Time-structure based reconstruction of physiological independent sources extracted from noisy abdominal phonograms *IEEE Trans. Biomed. Eng.* **57** 2322–30

Kanjilal P P and Palit S 1994 Extraction of multiple periodic waveforms from noisy data *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* pp II361–4

Kristjansson T, Attias H and Hershey J 2004 Single microphone source separation using high resolution signal reconstruction *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* pp II817–20

Paiva R P, Carvalho P, Henriques J, Antunes M, Muehlsteff J and Aubert X 2009 Assessing PEP and LVET from heart sounds: algorithms and evaluation *Proc. 31st IEEE EMBS Annu. Int. Conf.* pp 3129–33

Park T H 2000 Salient feature extraction of musical instrument signal *Masters Thesis* Dartmouth College

Paul A S, Wan E A and Nelson A T 2006 Noise reduction for heart sounds using a modified minimum-mean squared error estimator with ECG gating *Proc. 28th IEEE EMBS Annu. Int. Conf.* pp 3385–90

Potamitis I and Ozerov A 2008 Single channel source separation using static and dynamic features in the power domain *Proc. 16th European Signal Processing Conf. (EUSIPCO)* pp 1–5

Rajan S, Doraiswami R, Stevenson R and Watrous R 1998 Wavelet based bank of correlators approach for phonocardiogram signal classification *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis* pp 77–80

Schmidt S E, Holst-Hansen C, Graff C, Toft E and Struijk J J 2010 Segmentation of heart sound recordings by a duration-dependent hidden Markov model *J. Physiol. Meas.* **31** 513–29

Tavel M E 1967 *Clinical Phonocardiography and External Pulse Recording* (Chicago, IL: Year Book Medical Publishers)

Tsalaile T 2008 Digital signal processing algorithms and techniques for the enhancement of lung sound measurements *PhD Thesis* Loughborough University

Zhang Y and Zhang C 2006 Separation of music signals by harmonic structure modeling *Proc. Neural Information Processing Systems (NIPS)* pp 1617–24