35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Towards a Time-Feature Independent Phonocardiogram Segmentation

João P. Ramos*, Paulo Carvalho† and Miguel Coimbra‡

Abstract— **Delimitation and classification of each heart sound is a rather difficult task. Elevated heart rates, as found in pediatrics and in some adults as well, influence some of the most reliable features used by existing methods. Furthermore, in real life scenarios, cardiologists will not have the time to acquire the signal's length required by some of the existing algorithms, which make us think that different approaches ought to be pursued. This paper presents the work on heart sound segmentation using structural and energy based features. It is an attempt to not rely on features considered crucial to most existing approaches. Yet, it achieves a high sensitivity and specificity comparable to some literature.**

I. INTRODUCTION

The latest numbers show that cardiovascular diseases (CVD) are responsible for 47% of all deaths in Europe [1]. By the time an individual presents oneself at the hospital with CVD symptoms, it is typically at an advanced stage of the disease, which leads to a not very encouraging prognosis involving expensive treatments.

A more pro-active approach involving cheap cardiac health screening of the general population can help the physician detect possible complications at an early stage. Currently, two effective cardiac screening methodologies are the electrocardiogram (ECG) and echocardiograms exams but these can be expensive for mass screening and require technical expertise that is not available to most health professionals. Auscultation can be a powerful alternative to an ultrasound, providing a cheap and simple methodology for assessing the mechanical performance of the heart [2].

So far, heart sounds' analysis has been of great help in CVD diagnosis [3]. It is an inexpensive exam, the stethoscope is easy to carry and it can provide the physician discriminative information. Nonetheless, a detailed analysis of heart sounds requires a highly proficient cardiologist. With the development of the electronic stethoscope, computer aided auscultation systems have been built in order to process recorded signals and provide decision support to the less specialized physicians.

Segmentation of the phonocardiogram (PCG) onto the different phases of a cardiac cycle is of the uttermost importance. In a cardiac cycle we ought to find, at least, the first and second heart sound - S1 and S2, respectively. Each of these heart sounds can later be processed and meaningful features extracted to support a clinician's decision. However such a clean detection is hard to achieve due to the highly uncontrolled noisy environments of health centers such as hospitals and clinics. Furthermore, the discrimination between S1 and S2 is made more difficult due to the similarity of some of their characteristics.

One of the most robust techniques to perform heart sound segmentation is using ECG gating, as in [4], [5]. There is a direct relationship between ECG waves and the heart sound's main components. Nevertheless, this approach requires a synchronized ECG as reference, which is neither cheap nor practical at the moment for a mass cardiac screening objective. Therefore, much effort has been made to perform heart sound segmentation without a reference signal. The initial identification of possible candidates for the first and second heart sounds usually employs some sort of signal's envelogram, like the averaged Shannon energy in [6], [7], or homomorphic filtering [8], [9]. Afterwards, peaks are detected and classified into S1 or S2. In this classification process, it is standard to believe in physiologically inspired criteria such as saying the diastolic period is longer than the systolic, and that the systolic period remains reasonably constant throughout the acquisition. This kind of features usually involves requirements regarding the length of the acquisition in order to estimate heart rates, compare different systolic periods, etc. However, as reported in [10], an increased heart rate influences those intervals. This is common in the pediatric population and in some adult subjects. Moreover, the presence of noise artifacts, that may resemble a heart sound, has an effect on the whole classification. In such cases, the use of these type of time intervals as features is not viable.

This work's starting point is the method proposed in [7], [11], with which we find S1 and S2 candidates. However, when it comes to the classification of said candidates, to the extent of our knowledge, none of the work in the literature is able to perform segmentation without relying on some sort of physiological criteria based on systolic and diastolic periods.

Moreover, the introduction of the Teager energy operator as feature on the phonocardiogram seems to be a relatively recent subject. In fact, only recently have Fang et al [12] used the Teager energy operator as a heart rate estimator. As it seems, Teager energy has noteworthy characteristics that makes it of interest when processing phonocardiograms.

The following section provides a detailed description of the method proposed. Section III shows the results of the algorithm given a publicly available dataset. Finally, section IV concludes this paper with some final remarks and future

* João P. Ramos is with Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal and with the Faculty of Science and Technology, University of Coimbra, Portugal. jpramos @ dei.uc.pt

† Paulo Carvalho is with the Faculty of Science and Technology, University of Coimbra, Portugal. pcarvalho @ dei.uc.pt

‡Miguel Coimbra is with Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal. mcoimbra @ dcc.fc.up.pt

work.

II. METHODS

The methods presented here are divided into three modules. First the heart sound is pre-processed and the possible candidates for S1 or S2 are selected. Secondly, features for each of the selected candidates are extracted. Finally, the candidates are classified as one of the possible outcomes (Figure 1).

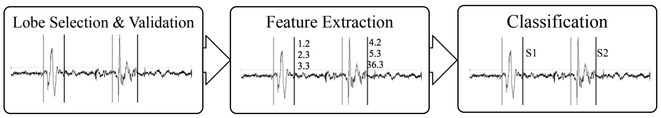

Fig. 1. Block diagram of the proposed algorithm.

A. Lobe selection

The lobe selection stage locates the heart sound, its boundaries and disposes non-relevant lobes such as noise or other artifacts' fluctuations. The basis of our lobe selection stage are the methods found in [7], [11]. Here, instead of the discrete wavelet decomposition, we perform a stationary wavelet transform decomposition. We found that by using the latter we obtain more precise boundaries, with no time-shifts or delays.

To find the best decomposition level, one ought to find where most of the frequencies of the signal under study reside. In [11], the authors define normal heart sounds as below $600Hz$, although the composition of the studied population is not clear. In our study, we find that our best results are using the $3^{rd} - 0Hz$ to $500Hz-$ decomposition level, which corroborates the above assumption.

To extract the envelope, an averaged Shannon energy (eq. 1) was applied on the approximation coefficients. The boundaries for each envelope's peak are found by first normalizing the envelope to mean 0 and standard deviation of 1, and afterwards whenever the normalized envelope zero-crosses, we mark the point as a candidate's delimitation.

At this stage, we have more candidates than actual heart sounds and some pruning is necessary. We have used some of the rules created in [7] to validate the candidates. Specifically, we have:

- The duration of a heart sound is no more and no less than $200ms$ and $25ms$, respectively;
- An interval between two heart sound candidates less than $50ms$ may be a S2 split

Afterwards, each validated candidate moves to the feature extraction stage.

B. Feature extraction

The features extracted are mostly energy based. Like in [7], we are looking for the high frequency markers found in S2. We have considered the Averaged Shannon Energy (ASE) and Teager Energy. The former is a well established PCG's envelope extraction method, whereas the latter has been used mostly in speech processing but because of the aforementioned characteristics it seems reasonable to explore its capabilities in PCG analysis.

The advantages of ASE are already explained in [6]. As in [6], here we have computed ASE with a moving window of .02 second and a .01 second of overlap (eq. 1).

$$ASE = -{}^{1}/_{N} \sum_{i=1}^{N} x_i^2 log_2(x_i^2) \quad (1)$$

where N is the length of the moving window and x is the signal.

The other energy calculated is the Teager Energy (TE). It is a non-linear measurement that can express the complexity of the signal. As shown in [13], TE substantially magnifies the high intensity parts of the signal. In fact, TE responds quadratically to changes in the amplitude and frequency of the signal [14]. Given the small differences, both in amplitude and frequency, between S1 and S2, this quadratic response can set the heart sounds apart. This, and the fact that it only needs three samples and is extremely easy to implement (eq. 2), makes it an extremely valuable asset.

$$TE(n) = x_n^2 - x_{n+1}x_{n-1} \quad (2)$$

where n is the sample being calculated, and x the signal under study.

The only feature not based on the energy of the signal compares the amplitude and width of each heart sound. The reasoning behind this feature lies on the observation that S2 has a larger amplitude than S1. The ratio between the amplitude and width of each heart sound is computed.

A full list, with a brief explanation of each feature, is provided:

- **TE_DETC** - Teager energy is applied to the details coefficients on the $3^{rd} - 500Hz$ to $1000Hz-$ decomposition level (Figure 2);
- **SH_TE_DETC** - An averaged Shannon energy is applied to **TE_DETC** to smooth non-cardiac bursts to which the TE is rather sensitive (Figure 3);
- **TE_SGN_FILT** - The signal is filtered with a 4^{th} order low-pass Butterworth and a cut-off frequency of $750Hz$. Afterwards, Teager energy is applied to the resulting signal (Figure 4);
- **ASP_RATIO** - Ratio between the amplitude and width of the heart sound;
- **SH_APPC** - The averaged Shannon energy envelope applied on the approximation coefficients of the 3^{rd} decomposition level, determined to estimate the lobes candidates, is used as feature as well (Figure 5);

Each candidate has its energy based features calculated by eq. 3:

$$En\_Feature = {}^{1}/_{CL} \sum_{i}^{CL} En(i)^2 \quad (3)$$

where CL is the length of the heart sound candidate and En is one of each of the energy based features listed above.

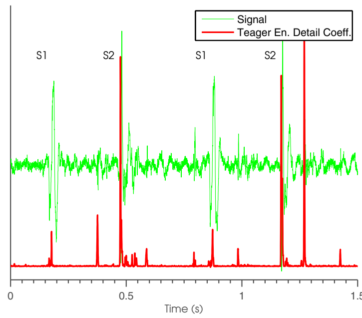

Fig. 2. Teager energy applied on the details coefficients of the 3^{rd} decomposition level (**TE_DETC**).

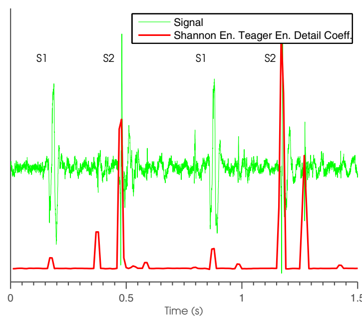

Fig. 3. ASE is applied to **TE_DETC** to smooth non-cardiac bursts (**SH_TE_DETC**).

C. *Classification*

To tackle the classification problem we group the candidates in pairs and use a simple feature comparison system. Here we assume that every S1 is followed by a S2 or vice-versa. The rules are the following:

- $RULE_1$ - $TE\_DETC_{S1} < TE\_DETC_{S2}$;
- $RULE_2$ - $SH\_TE\_DETC_{S1} < SH\_TE\_DETC_{S2}$;
- $RULE_3$ - $TE\_SGN\_FILT_{S1} < TE\_SGN\_FILT_{S2}$;
- $RULE_4$ - $ASP\_RATIO_{S1} < ASP\_RATIO_{S2}$;
- $RULE_5$ - $SH\_APPC_{S1} \geq SH\_APPC_{S2}$;

The first three rules exploit the high frequencies found in S2 (as seen in figures 2, 3 and 4). As aforementioned, Teager energy highlights this difference quite well with its quadratic response of the frequency and amplitude of the signal. The fourth rule tries to capture the difference in

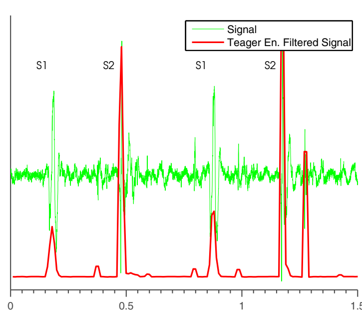

Fig. 4. Teager energy applied on the filtered signal (**TE_SGN_FILT**).

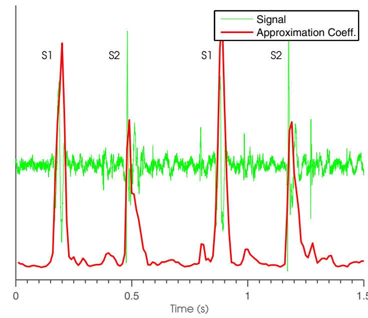

Fig. 5. ASE envelope applied on the approximation coefficients of the 3^{rd} decomposition level (**SH_APPC**).

amplitude between the two heart sounds. In all four cases, we expect higher values in the S2 heart sound. Lastly, the fifth rule is the averaged Shannon energy applied on the approximation coefficients of the $3^{rd} - 0Hz$ to $500Hz-$ decomposition level, which means that since S1 has lower frequency components, they will be portrayed as more prominent (as observed in figure 5). In this case, one will be looking for higher values in the S1 heart sound.

Following the diagram provided in figure 6, we can observe each of the rules used on the top. The output of a rule is binary, hence the $[0|1]$ as in a regular expression alike format. Each rule votes, with an equal weight, and the majority decides whether the pair under scrutiny is an S1 followed by an S2 or vice-versa.

By grouping and classifying the candidates in pairs, we are confining any transient noise or other artifact's influence to that specific location, letting the rest of the classification unscathed.

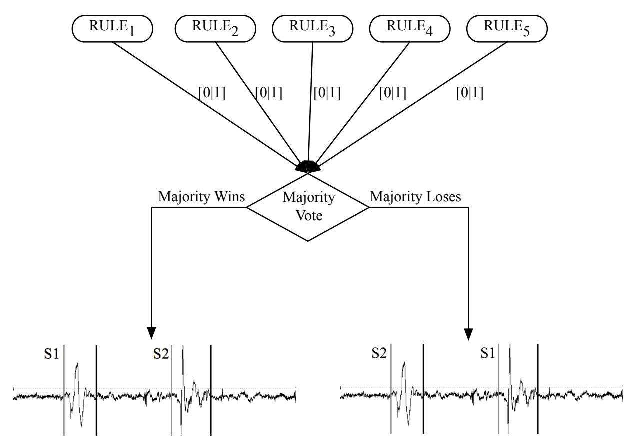

Fig. 6. Majority voting process used to classify a pair of heart sounds.

III. Results and Discussion

The performance of the proposed algorithm was assessed using the publicly available PASCAL CHSC2011 dataset [15]. Specifically, we have used samples from the dataset B originated from clinical trials in hospitals using the digital stethoscope platform DigiScope[1]. This dataset is mainly

[1]http://digiscope.up.pt/

composed by acquisitions performed in a pediatric population and with periods that vary between 1 and 30 seconds and a sampling frequency of $4000Hz$. We have not used the dataset in its entirety since some of the files were not adequately annotated. Nevertheless, it has the characteristics of the dataset's population and the samples' length that pose as challenges to most state of the art segmentation algorithms.

Since the annotation is a temporal location of the heart sound and we are actually interested in identifying not only its location but also the boundaries of the heart sound, we have considered a correct classification in those cases where the annotated point is in between of the limits found.

The results are summarized in Table I where the proposed algorithm is compared with a state of the art algorithm, [6].

Unlike the proposed algorithm, [6] performs a global classification, i.e. the classification is done after the discovery of the longest diastolic period, and this has repercussions on the final classification whenever the signal does not follow the common criteria. Aside from the sensitivity and specificity values it is worth noting the difference in the number of not classified heart sounds, where 73 heart sounds are not classified because of the segmentation method used in [6]. Moreover, on the proposed algorithm, 13 out of the 14 not classified heart sounds only exist due to the odd number of heart sounds in some samples and the fact that the classification is processed in pairs. So, in reality, only one annotation was not classified.

	Sensitivity	Specificity	Not Classified / Total
Proposed	92.8%	92.6%	$14/850$
Liang's [6]	77.1%	78.3%	$73/850$

TABLE I

CLASSIFICATION'S RESULTS.

IV. CONCLUSIONS

Several segmentation methods for heart sound have been proposed in the last decades. Nonetheless, the fact that the heart sounds have such frail boundaries when it comes to characteristics makes the task of discrimination between different heart sounds rather difficult.

This work is an attempt to take a different approach on the problem at hand and explore a set of features more focused on structural properties of the heart sounds rather than the periods of time that separate them.

The results, so far, have shown that this approach may be quite promising. We have found out that Teager energy provides reliable information regarding the qualities of both heart sounds and as such is an useful asset in heart sound segmentation. We wonder if this reliability is extended to further steps in heart sound processing.

Given the simplicity of the algorithm, there is plenty of room for improvement. Although we have mitigated the effect of transient noises to their specific location, given the classification approach used, they still remain a concern. The inclusion of other features could add robustness to the algorithm and starting to consider heart murmurs in the segmentation process would turn the algorithm into something more versatile.

ACKNOWLEDGMENT

This work was supported by: Instituto de Telecomunicações: DECABio (IT/LA/01075/2011); and FCT (Portuguese Science and Technology Agency) grants: PTDC/EIA-CCO/100844/2008 (DigiScope) and PTDC/EEI-PRO/2857/2012 (HeartSafe).

REFERENCES

[1] M. Nichols, N. Townsend, R. Luengo-Fernandez, J. Leal, P. Scarborough, and M. Rayner, *European Cardiovascular Disease Statistics 2012*. European Heart Network, Brussels, European Society of Cardiology, Sophia Antipolis, 2012.

[2] C. Shub, "Echocardiography or auscultation? how to evaluate systolic murmurs." *Canadian family physician*, vol. 49, no. 2, pp. 163–167, 2003.

[3] M. Tavel, "Cardiac auscultation," *Circulation*, vol. 113, no. 9, pp. 1255–1259, 2006.

[4] A. Quiceno, E. Delgado, M. Vallverd, A. Matijasevic, and G. Castellanos-Domnguez, "Effective phonocardiogram segmentation using nonlinear dynamic analysis and high-frequency decomposition," in *Computers in Cardiology, 2008*. IEEE, 2008, pp. 161–164.

[5] A. Gharehbaghi, T. Dutoit, A. Sepehri, P. Hult, and P. Ask, "An automatic tool for pediatric heart sounds segmentation," in *Computing in Cardiology, 2011*. IEEE, 2011, pp. 37–40.

[6] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*. IEEE, 1997, pp. 105–108.

[7] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, L. Eugenio, R. Schmidt, and J. Habetha, "Detection of s1 and s2 heart sounds by high frequency signatures," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 1410–1416.

[8] D. Gill, N. Gavrieli, and N. Intrator, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," in *Computers in Cardiology, 2005*. IEEE, 2005, pp. 957–960.

[9] C. Gupta, R. Palaniappan, and S. Swaminathan, "Classification of homomorphic segmented phonocardiogram signals using grow and learn network," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006, pp. 4251–4254.

[10] T. Bombardini, V. Gemignani, E. Bianchini, L. Venneri, C. Petersen, E. Pasanisi, L. Pratali, D. Alonso-Rodriguez, M. Pianelli, F. Faita, *et al.*, "Diastolic time–frequency relation in the stress echo lab: filling timing and flow at different heart rates," *Cardiovascular ultrasound*, vol. 6, no. 1, p. 15, 2008.

[11] H. Liang, S. Lukkarinen, and I. Hartimo, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 4. IEEE, 1997, pp. 1630–1633.

[12] Y. Fang, H. Wang, S. Guo, Q. Zhang, W. Liu, X. Wu, and X. Xu, "A novel cardiac cycle detection algorithm based on matched filter?" *Journal of Computational Information Systems*, vol. 8, no. 10, pp. 4135–4142, 2012.

[13] J. Jasper and K. Othman, "Feature extraction for human identification based on envelogram signal analysis of cardiac sounds in time-frequency domain," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 2. IEEE, 2010, pp. V2–228.

[14] J. Kaiser, "On a simple algorithm to calculate the energy of a signal," in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. IEEE, 1990, pp. 381–384.

[15] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor, "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results," http://www.peterjbentley.com/heartchallenge/index.html.