A Study of Heart Sound Analysis Techniques for Embedded-Link e-Health Applications

Hao-Dong Yao, Jia-Li Ma, and Ming-Chui Dong

***Abstract*—** Auscultation of heart sound (HS) signals serves as an important primary approach to diagnose cardiovascular diseases (CVDs) for centuries. Tackling the drawbacks of traditional auscultation with intrinsic restriction of ears limitation and low efficiency, automatic auscultation using embedded-link devices is able to provide timely intelligent HS interpretation and diagnosis for anyone, anytime, and anywhere conveniently. To explore a HS analysis method with simplicity, high efficiency, rapidity and convenience for resource-limited embedded-link e-health apparatus, this paper conducts a synthetic investigation of existing prevalent and up-to-date HS processing techniques, including HS de-noising and analysis. HS analysis approaches are generally categorized into feature-based analysis and entire-HS analysis without demanding feature extraction. After thorough study and comparison, the feature-based HS analysis method using discrete wavelet transform (DWT) is discovered as the best candidate for embedded-link e-health applications due to its high performance, good robustness and rapid computation.

***Keywords*—**automatic asusculation, discrete wavelet transform, feature extraction, signal analysis

I. Introduction

FOR centuries, cardiovascular diseases (CVDs) remain the leading cause of death throughout the world [1]. Auscultation of heart sound (HS) signals plays a vigorous role in CVDs early prevention and detection. By directly listening to HS with an acoustic stethoscope, it provides a cost-effective approach to inspect abnormality of HS signals having potential pathological CVDs symptoms. Unfortunately, the intrinsic disadvantages of traditional HS auscultation have significantly hindered its profound development, such as: (1) inherent restriction due to human ears limitation; (2) inefficiency in qualitative and quantitative analysis; (3) insufficiency in tackling with inaudible HS components; (4) disability to record, replay, and store the HS signals; (5) subjectivity of the analyst with spotty training and experiences. As a solution, automatic auscultation using electronic stethoscope has been enjoying a booming development recently. Furthermore, for portable usability, an intelligent e-health auscultation system resided on embedded-link mobile devices is grossly attractive to provide timely intelligent HS interpretation and diagnosis for anyone, anytime, and anywhere conveniently.

Research supported by the Research Committee of University of Macau under Grant MYRG184 (Y2-L3)-FST11-DMC, and Science and Technology Development Fund (FDCT) of Macau under Grant 018/2009/A1.

Hao-Dong Yao, Jia-Li Ma, and Ming-Chui Dong are with Electrical and Computer Engineering Department, Faculty of Science and Technology, University of Macau, Macau S.A.R., China. (email: yhd1992@126.com; jimohanqiu@hotmail.com; mcdong@umac.mo).

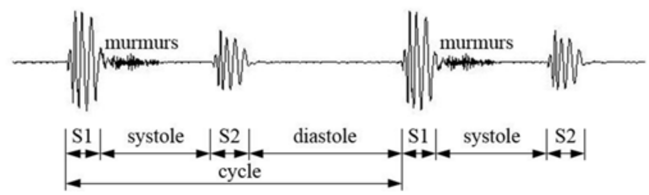

Fig. 1 Waveform of a representative HS signal.

For embedded-link e-health applications, the HS analysis approaches should meet the following technical requirements: (1) simplicity due to the resource limitation and computation ability of embedded-link mobile devices; (2) rapidity to provide quasi-real-time diagnostic results for on-site users; (3) high efficient to handle diversified HS signals and diseases; (4) convenience with provision of local diagnosis independence of server or hospital doctors. During the past decades, various HS analysis techniques have been exploited [17], [18], [20], [24]. Our research group commits great efforts on it too [2], [3], [24]. The targets of this paper are to summarize the formidable bottleneck problems and reveal the relevant modern technologies, seek an appropriate HS analysis approach for embedded-link e-health applications through exhaustively reviewing prevalent and up-to-date HS analysis techniques as well as based on our group's R&D experiences.

II. HS Signals

A. HS Signals Description

Heart sound, or phonocardiogram (PCG), is the repetitive "lub-dub" sound generated by heartbeats and its representative waveform is shown in Fig. 1. Three conceptions are related with signals: (1) fundamental heart sounds (FHS), including S1 and S2 caused by the sudden closure of the mitral and tricuspid valves, and the closure of aortic and pulmonary valves respectively [4]; (2) abnormal heart sounds (AHS), including S3 and S4 (always inaudible) generated by the rapid ventricular filling in early diastole, and the ventricular filling due to atrial contraction separately; (3) murmurs, divided as innocent murmurs and harmful murmurs, produced by turbulent blood flow through a blocked valve or backward flow through a leaking valve. FHS, AHS, and murmurs are all important indicators of CVDs. Besides these components, several HS activities are also utilized in CVDs diagnosis, including systole (defined as the interval between the end of S1 to the start of

http://dx.doi.org/10.15242/IIE.E0414082

same cycle's S2) and diastole (defined as the interval between the end of S2 to the start of next cycle's S1) as depicted in Fig. 1.

B. HS Acquisition

In HS automatic auscultation, electronic stethoscope is used to acquire and record the HS signals efficiently. Since the frequencies of all important HS components are below 1000 Hz, an appropriate sampling frequency varying from 4000 to 20000 Hz is usually employed. It is noteworthy that the correct position of auscultation is very important for acquiring qualified signals due to the fact that different CVDs should sample HS from different position of chest.

C. HS Benchmark Databases

The HS benchmark database is essential for HS study and tests. Table I lists several popular HS benchmark databases, including qdheart [5], eGeneral Medical (eGM) [6], cardiac auscultatory recording database (CARD) [7], and Thinklabs [8]. Among these databases, qdheart had been mostly used until 2010, but not available since then. Although eGM is also commonly used, the provided signals are fake and artificially generated. The signals provided by CARD are original HS signals directly acquired from patients at Johns Hopkins Hospital. In Table I we discover and show it out that, the parameters of listed databases differ a lot which leads to incomparable results among methods using different databases. Thus, a standard HS benchmark database with uniform settings, parameters, and doctor's indices under standard sampling criteria is indispensable.

III. HS De-Noising

The HS signals are nonstationary and imperfect blended with noises from sounds of breathes, contact of stethoscope with chest skin and other ambient sounds. Therefore, to obtain the qualified HS signals for further analysis and interpretation, HS de-noising is crucial mission in signal pre-processing. Here two most appreciated de-noising technologies are introduced.

A. Short-Time Fourier Transform

The principle of short-time Fourier transform (STFT) is to acquire the time-frequency (TF) characteristics of HS with a short sliding time window. The time window should be short enough to guarantee the signal stationary within the short period. By conducting Fourier transform on the windowed signals, the time-varying frequencies results could be obtained. Since the windowed HS components are quasi-stationary yet the noises are not, it could be distinctly observed that in TF domain the TF results of HS components congregate while the noises are dispersed. Consequently the noises could be separated from HS components through fuzzy detection successfully. However, for signals with sudden changes, it is troublesome to find an appropriate time window length since the shorter window could guarantee the signal stationary but at the cost of reduction in frequency resolution [9].

B. Discrete Wavelet Transform

Different from STFT, discrete wavelet transform (DWT) uses variable window sizes thus able to obtain good time resolution at high frequency and good frequency resolution at low frequency. DWT decomposes windowed signals into shifted and scaled version of the mother wavelet. Studies show that after DWT decomposition, the obtained wavelet coefficients of HS signals are greatly larger than coefficients of noises [10]. Therefore, decomposition items with coefficients below a certain level are discarded as noises and the remained decomposition items are constructed to obtain noise free HS signals.

IV. HS Analysis

HS analysis aims to map the input HS signals into the disease categories, which could be double classes ("healthy" and "diseased") or multiple classes with detailed type of disease. Two kinds of HS analysis methods are commonly used, feature-based HS analysis method and entire-HS analysis method without demanding feature extraction.

A. Feature-Based HS Analysis

Features are defined as specific parameters or characteristics extracted from the acquired signals that have potential to discriminate HS classes [11]. Normally three generic steps are contained in feature-based HS analysis: segmentation, feature extraction, and classification.

1. Segmentation

Segmentation is to identify the boundaries of cardiac cycles and HS components from contiguous HS signals. HS segmentation can be divided into two manners. One segmentation manner is based on electrocardiology (ECG) signal. Regarding ECG as the reference signal, the locations of

TABLE I
HS Benchmark Databases

Database	Sampling Frequency (Hz)	Resolution (bits)	Number of Recordings	Provider
qdheart [5]	22050	16	42	The Affiliated Caediovascular Hospital of Medical College QingDao University
eGM [6]	5000, 8000, 8012, 11025, 22050, 22257	N/A	64	eGeneral Medical Inc., USA
CARD [7]	4000	16	N/A	Johns Hopkins University
Thinklabs [8]	N/A	16	22	Thinklabs

S1 and S2 can be traced according to the QRS wave and T wave respectively [12]. Unfortunately, this manner requires extra acquisition, storage, and analysis of ECG signal, which increases the computing burden and generally is unsuitable for portable usability. Another segmentation manner is based on envelope analysis. After obtaining the envelope of HS signal, the envelop peaks are detected which correspond to HS peaks. Consequently, the cardiac cycles are segmented. References [13]-[16] are examples of this manner with various envelop detection methods. Reference [13] employs the energy distribution of HS signals to form the envelope, and then a threshold value is set to detect the peaks of S1 and S2. Afterwards, the systole and diastole are recognized according to the time duration characteristics. However, unqualified HS signals or inappropriate threshold values induce incorrect segmentation results. Reference [14] is based on homomorphic envelogram, in which by means of homomorphic filtering, the smooth HS signal envelogram is acquired to detect successfully the HS components. Even for signals with low peaks, artifacts, or signal splits, this method performs segmentation robustly. Reference [15] adopts DWT to remove the murmur first, thus to eliminate its effects on segmentation accuracy, afterwards, the envelope is calculated for segmentation [16].

2. Feature Extraction

Feature extraction is to calculate the identifying parameters or characteristics from each cycle. Based on the variety of utilized characteristics, HS feature extraction approaches are roughly classified into acoustic feature extraction and TF feature extraction.

2.1) Acoustic Feature Extraction

Acoustic feature extraction methods are derived from the human auditory perception systems in tradition auscultation diagnosis. Two typical acoustic feature extraction methods are Mel-frequency cepstral coefficient (MFCC) and timbre extraction.

MFCC is based on the theory of human auditory system that human audition spaces linearly at low frequency band and logarithmically at high frequency [17]. First, discrete Fourier transform (DFT) is conducted on the windowed overlapping HS segments to get the energy distribution in frequency domain. After that, a set of triangle filters are adopted to get the logarithmic Mel spectrum as the inputs of discrete cosines transform (DCT). Finally, the cepstral coefficients are obtained and utilized as the HS features for further analysis. Since it simulates the human ear sound processing system, MFCC is keen at low frequency band where HS signals locate and exhibits robustness even with noisy signals [18]. However, the two transforms contained in MFCC algorithm lead to an increase in computation complexity and time.

The foundational principle of timbre extraction is that with different CVDs HS signals possess different timbre characteristics which provide the reference for CVDs recognition. The timbre analysis algorithm in [19] is suggested to differentiate instruments with different timbres initially, and thus extract timbre features based on the descriptors of MPEG7 (moving picture expert group 7) such as harmonic centroid, log attack time, spectral centroid, and temporal centroid. For better results, some additional descriptors like tristimulus parameters and transient duration could also be used. Compared with MFCC, this method is much easier and saves lots of computation. Our team is attempting to apply principle of this method to extract timbre features of HS signals instead of instruments, and it is still under development so far.

2.2) TF Feature Extraction

Since HS signals are nonstationary with marked changes in time and frequency, only time or frequency features are insufficient to support the diagnosis. Hence, TF features reflecting HS pathological information in both dimensions that may not be heard or seen in the raw HS or HS waveform will contribute to high accuracy in the following classification. The principle of TF is to extract the time and frequency features simultaneously using various transformations, such as STFT, and DWT.

It is mentioned above that STFT provides TF features by using a short sliding window. Since signals with different diseases exhibit different TF distributions [20], thus the achieved TF features are potential to be used in signals recognition and CVDs diagnosis. However, STFT cannot track sensitive sudden changes in time domain and suffers from the tradeoff between time and frequency resolution.

DWT was developed as a method to obtain high-resolution time and frequency information simultaneously. With the aid of low-pass and high-pass filters, the input signal is decomposed into subband signals as approximations and details. These subband signals are much more distinct to exhibit HS components, which is beneficial for extracting HS features such as the rhythm and intensity of HS components. It is worthwhile to note that the widths of filter banks are varied with the decomposition levels. While a wide filter gives rise to high frequency components, a narrow filter picks up the low frequency components. Consequently, the obtained multiresolution components could overcome the problem of TF resolution tradeoff as in STFT. In addition, DWT also leads to a considerable reduction in computing time [20].

3. Classification

Classification is to categorize the nature of HS with the aid of extracted parameters and characteristics. Several prevalent classification methods including artificial neural networks (ANN), hidden Markov models (HMM), support vector machine (SVM), and dynamic time warping (DTW) are summarized here.

ANN is a highly interconnected system of computational nodes or neurons. A typical example of ANN is back-propagation neural network (BPNN) which generally consists of input layer, hidden layers, and output layer [21]. ANN system is optimized through the iteratively training procedure and the outputs converge to the training data. Due to its simple structure and flexibility for solving nonlinear complex problems, ANN is widely used for dealing with multiple dimensions and continuous features classification. However, long training time and large training sample set are required.

HMM is a probabilistic state machine in which the states of the machine are unobservable, yet the outputs are observable. HMM with its Markov chain structure can inherently incorporate the time sequential character of the signal. By using the Gaussian mixture densities, the HMM is also expected to faithfully represent the various spectral characteristics of the HS signal. In a recent study [22], it is found that HMM performs much better than ANN in classifying HS signals with 10 different diseases.

SVM is a method of machine learning based on statistical theorems. It could simultaneously minimize the empirical classification error and maximize the geometric margin. However, a large sample set is required in order to guarantee its prediction accuracy.

DTW algorithm is based on the idea of dynamic optimization in order to find the minimum characteristic values of the time calibration path between reference signal and test signal [23]. DTW is prevalently used in signal processing because it can efficiently minimize the effects of shifting and distortion for time sequences. Besides, it is capable of matching two sequences of unequal length and has strong anti-white noises ability [23].

B. *Entire-HS Analysis*

In entire-HS analysis methods, the HS are processed and classified as a whole directly without the need of feature extraction. The complexity and similarity analysis (CSA) in [24] is an example of entire-HS analysis methods.

In CSA, the HS signals are treated as a whole, avoiding the operations of segmentation and feature extraction. First, the N-gram code of the input HS signal is calculated using musical instrument digital interface (MIDI) coding and ASCII coding. After that, the Lempel-Ziv (LZ) complexity between N-gram of input signal with unknown disease and the N-gram of the reference signals with known diseases in database is computed one by one. From the complexity results, similarity score between them can be obtained. With a higher similarity between input signal and certain reference signal, it is more convincing to diagnose the input signal as the corresponding disease.

Since it releases the requirements of segmentation and feature extraction, CSA is robust to freaky signals and free from segmentation errors. However, it suffers numerous computations which restricts its wide applications.

V. Discussion

A comparison between several prevalent and up-to-date works is made as shown in Table II. The employed benchmark databases, algorithms, as well as the performances are listed for an intuitive description. It is noteworthy that when interpreting the results shown in Table II, a direct comparison based only on the numerical results is misleading, since they are taken in various test conditions with differed samples. Nevertheless, this table remains useful for offering a general idea of listed techniques.

As shown in Table II, CSA is one kind of entire-HS analysis method while others are all based on feature extraction. CSA method obtains a diagnosis accuracy of 74.8% for handling 9 types of CVDs which is the lowest compared with other methods. In addition, the numerous encoding procedures in CSA algorithm induce a large computation quantity. In Table II, three MFCC-based approaches are also listed. Although the achieved accuracy is higher than 80%, the MFCC-based methods suffer from computation complexity caused by two inverse transforms. DWT method exhibits advantageous diagnosis performance with accuracy as high as 99% for recognizing 7 types of CVDs. DWT also has superior robustness for noisy HS signals since the diagnosis accuracy is 90% under 10 dB white noise test and 92% without noises. Furthermore, DWT is outstanding in HS signals de-noising and segmentation. Segmentation is even nonessential in DWT methods so that the

TABLE II
Comparison Between Different HS Analysis Methods

Method	HS Data	Need of Segmentation	Need of Feature Extraction	Classification Technique	Diagnosis Accuracy	Classification Types	Remarks
CSA	eGM, Cadionics [26]	No	No	Inference Machine	74.8% [24]	Multiple classes (Normal and 9 CVDs)	Complex Encoding
MFCC	Site-sampled	Yes	Yes	BPNN	80% [18]	Double classes (Normal and Diseased)	Complex Computation
	Databese [29]	Yes	Yes	DTW	92.5% without noises; 91.6% with 40dB white noise [23]	Multiple classes (Normal and 5 CVDs)	
	Site-sampled, database [27]	Yes	Yes	HMM	99.21% [1]	Multiple classes (Normal and 9 CVDs)	
DWT	Site-sampled	Yes	Yes	Grow and Learn Networks	99% [25]	Multiple classes (Normal and 6 CVDs)	Short Computing Time
	Site-sampled	Yes	Yes	Linear Vector Quantization Networks	96% [25]	Multiple classes (Normal and 6 CVDs)	
	Site-sampled, database [28]-[31]	No	Yes	Neural Networks	92% without noises; 90% with10 dB white noise [16]	Double classes (Normal and Diseased)	

segmentation errors can be avoided. In conclusion, regarding the requirements of simplicity, rapidity, high efficiency, and convenience, DWT method is discovered as the best candidate for HS analysis in embedded-link e-health applications.

VI. Conclusion

Aim at exploring a dedicated HS analysis approach for embedded-link e-health applications, this paper conducts a thorough reviewing of the existing prevalent and up-to-date HS analysis techniques. After study in depth and comprehensive comparison, the feature-based HS analysis method using DWT is the preferred option due to its high performance, good robustness, and short computing time.

The future work focuses on adapting DWT methods for HS analysis and fulfilling the implementation of automatic HS auscultation and diagnosis on embedded-link mobile devices.

References

[1] S. Chauhan, P. Wang, C. S. Lim, and V. Anantharaman, "A computer-aided MFCC-based HMM system for automatic auscultation," *Computers in Biology and Medicine*, vol. 38, pp. 221–233, Feb. 2008. http://dx.doi.org/10.1109/TBME.2010.2051225

[2] X. Zheng, M. Dong, B. Fu, and X. Fei, "Automatic pathological signal detection of heart sound based on multi-resolution analysis," in *Proc. Mathematics and Computers in Contemporary Science*, Milan, 2013, pp. 96–100.

[3] X. Zheng, B. Fu, B. D. Sekar, and M. Dong, "Complexity and similarity approach based on heart sound murmurs for cardiac pathological status analysis," in *Proc*. *Biologically Inspired Computation 2012(BIC 2012)*, Faro, Portugal, 2012, pp. 206–211.

[4] M. S. Manikandan, and S. Dandapat, "Wavelet energy based compression of phonocardiogram (PCG) signal for telecardiology," in *IET-UK International Conference on Information and Communication Technology in Electrical Sciences*, Tamil Nadu, India, 2007, pp. 650–654.

[5] http://www.qdheart.com/cgz/jcxt/tyvf/tyvf1.htm

[6] http://www.egeneralmedical.com/listohearmur.html

[7] C. Tuchinda, and W. R. Thompson, "Cardiac auscultatory recording database: delivering heart sounds through the Internet," in *Proc. the American Medical Information Association Symposium*, Washington, 2001, pp. 716–720.

[8] http://www.thinklabs.com/#!heart-sounds/c1dqs

[9] H. Tang, T. Li, Y. Park, and T Qiu, "Separation of heart sound signal from noise in joint cycle Frequency-Time-Frequency domains based on fuzzy detection," *IEEE Trans. Biomedical Engineering*, vol. 57, pp. 2438–2447, 2010. http://dx.doi.org/10.1109/TBME.2010.2051225

[10] S. R. Messer, J. Agzarian, and D. Abbott, "Optimal wavelet denoising for phonocardiograms," *Microelectronics Journal*, vol. 32, pp. 931–941, Dec. 2001. http://dx.doi.org/10.1016/S0026-2692(01)00095-7

[11] A. Cheema, and M. Singh, "Steps involved in heart sound analysis-a review of existing trends," *International Journal of Engineering Trends and Technology*, vol. 4, pp. 2921–2925, Jul. 2013.

[12] A. Iwata, N. Ishii, and N. Suzumura, "Algorithm for detecting the first and the second heart sounds by spectral tracking," *Medical and Biological Engineering and Computing*, vol. 18, pp. 19–26, Jan. 1980. http://dx.doi.org/10.1007/BF02442475

[13] S. Choi, and Z. Jiang, "Comparison of envelope extraction algorithms for cardiac sound signal segmentation," *Expert Systems with Applications*, vol. 34, pp. 1056–1069, Feb. 2008. http://dx.doi.org/10.1016/j.eswa.2006.12.015

[14] 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*, Lyon, 2005, pp. 957–960.

[15] L. Huiying, L. Sakari, and H. Iiro, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *Proc. the 19th Annual International Conference of Engineering the Medicine and Biology Society*, Chicago, 1997, pp. 1630–1633.

[16] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," *Biomedical engineering online*, 2011. http://dx.doi.org/10.1186/1475-925X-10-13

[17] P. Wang, Y. Kim, and C. B. Soh, "Feature extraction based on Mel-scaled wavelet transform for heart sound analysis," in *Proc. the 27th Annual Conference of Engineering in Medicine and Biology*, Shanghai, 2005, pp. 7572–7575.

[18] I. Kamarulafizam, S. H. Salleh, J.M. Najeb, A. K. Ariff, and A. Chowdhury, "Heart sound analysis using MFCC and time frequency distribution," in *World Congress on Medical Physics and Biomedical Engineering 2006*, Seoul, 2007, pp. 946–949.

[19] X. Zhang, and Z. W. Ras, "Analysis of sound features for music timbre recognition," in *2007 International Conference on Multimedia and Ubiquitous Engineering*, Seoul, 2007, pp. 3–8. http://dx.doi.org/10.1109/MUE.2007.85

[20] S. M. Debbal, and F. Bereksi-Reguig, "Computerized heart sounds analysis," *Computers in Biology and Medicine*, vol. 38, pp. 263–280, Feb. 2008. http://dx.doi.org/10.1016/j.compbiomed.2007.09.006

[21] C. G. DeGroff, S. Bhatikar, J. Hertzberg, R. Shandas, L. Valdes-Cruz, and R. L. Mahajan, "Artificial neural network–based method of screening heart murmurs in children," *Circulation*, vol. 103, pp. 2711–2716, Jun. 2001. http://dx.doi.org/10.1161/01.CIR.103.22.2711

[22] Y. J. Chung, "A classification approach for the heart sound signals using hidden markov models," in *SSPR/SPR*, Hong Kong, 2006, pp. 375–383.

[23] W. Fu, X. Yang, and Y. Wang, "Heart sound diagnosis based on DTW and MFCC," in *2010 3rd International Congress on Image and Signal Processing (CISP2010)*, Yantai, 2010, pp. 2920–2923.

[24] B. D. Sekar, J. Y. Dou, B. B. Fu, X. L. Fei, and M. C. Dong, "Complexity and similarity analysis of heart sound for cardiovascular disease detection," *Biotechnology*, vol. 10, pp. 316–322, Sep. 2011. http://dx.doi.org/10.3923/biotech.2011.316.322

[25] T. Ölmez, and Z. Dokur. "Classification of heart sounds using an artificial neural network," *Pattern Recognition Letters*, vol. 24, pp. 617–629, Jan. 2003. http://dx.doi.org/10.1016/S0167-8655(02)00281-7

[26] http://www.cardionics.com/learningcardiac.html

[27] http://www.bcm.tmc.edu/class2000/sims/HeartSounds.html

[28] http://www.dundee.ac.uk/medther/Cardiology/hsmur.html

[29] http://www.texasheart.org/Education/CME/index.cfm

[30] http://int-prop.lf2.cuni.cz/heart_sounds/h12/index.html

[31] http://sprojects.mmi.mcgill.ca/mvs/mvsteth.htm