Multidimensional Feature Extraction Based on Timbre Model for Heart Sound Analysis

Hai-Yang Wang, Guang-Pei Li, Bin-Bin Fu, Jun Huang, and Ming-Chui Dong

***Abstract*—Cardiovascular diseases (CVDs) are currently the leading cause of deaths worldwide. The traditional auscultation is cost-effective and time-saving for the public to diagnose CVDs early. While many approaches for analysis of the heart sound (HS) signal from auscultation have been utilized successfully, few studies are focused on acoustic perspective to interpret the HS signal. This paper creatively proposes a multidimensional feature extraction technique based on timbre model to interpret HS, which stems from clinical diagnostic basis and medical knowledge. The extracted features have three dimensions, including spectral centroid (SC), log attack time (LT) and temporal centroid (TC). The simulation experiments indicate that the proposed method is promising in HS feature extraction and the later CVD diagnosis.**

***Index Terms*—Acoustics, feature extraction, Hilbert transform, spectral centroid, temporal centroid**.

I. Introduction

Cardiovascular diseases (CVDs) have become a great threat to human' lives aggressively. A good way to prevent the death caused by CVDs is early discovery and interposition. Over the years, in spite of the advent of echocardiogram (ECHO), electrocardiogram (ECG), photoplethysmogram (PPG), sphygmogram (SPG), etc., the heart auscultation signal is still one of the most primary physiological signals. Physicians primarily analyze it thus to find signs of pathologic conditions, as it can provide clues to the diagnosis of many cardiac abnormalities including valvular heart disease, congestive heart failure and congenital heart lesions before requesting for ECHO, ECG, PPG, SPG, etc. Notably, in remote areas or less developed regions, auscultation may be the only means available. This leads the research on heart sound (HS) signal analysis and interpretation in order to provide a cost-effective and time-saving prognostic approach for the victims of CVDs.

As for the analysis of HS, many approaches have been proposed in feature exaction in the literatures such as wavelet decomposition and reconstruction method [1]-[3], short time Fourier transform (STFT) method [4], [5], and S-transform method [6], [7]. Most solely analyze the time frequency domain for feature extractions. This leads to a conclusion that feature extractions are less aligned with medical knowledge.

A doctor diagnoses the CVDs generally through a stethoscope from acoustic perspective. Common descriptive terms about what it sounds like in auscultation include rumbling, blowing, machinery, scratchy, harsh, gallop, ejection, click, drum, cymbal, etc. From acoustic perspective to analyze bio-signal, its main advantage is that engineering is aligned with medicine. For instance, continuous machinery sound is heard on the left sternal border between the second and third ribs in auscultation indicating the patent ductus arteriosus. From acoustic perspective, the mel-frequency cepstral coefficient (MFCC) method has been utilized for HS feature extractions [8]-[12]. MFCC is based on the theory that human audition spaces are linearly at low frequency band and logarithmically at high frequency. As existing two inverse-transforms in MFCC, it encounters computation complexity.

There exists another approach named timbre analysis from acoustic perspective, which hasn't been utilized and can avoid computation complexity. In acoustics, timbre is a significant attribute of three acoustic attributes, which embodies the texture of acoustic source. CVDs, as the pathological changes in the heart and the blood vessels, provide different timbre information. As a result, the timbre is well suitable for HS feature extractions. However, current timbre model analysis all aim at different music instrument recognition and little literature ever reported its exploration on HS feature extraction. In this paper, the timbre model based multidimensional feature extraction technique for HS analysis is creatively proposed. It is observed from simulations that the proposed method is capable of characterizing features with diagnostic significance from HS.

II. Methodology

Timbre expresses normally the quality of sound that distinguishes one music instrument from another among a wide variety of instrument families and individual categories. Virtually timbre is a subjective judgment and must be parameterized very carefully thus to apply it on automatic timbre recognition. Psychoacoustic literatures conclude that the dimensions of timbre can be categorized as fundamental frequency, log attack time, temporal centroid, spectral centroid, harmonics, etc. [12]-[17].

In this paper, spectral centroid (*SC*), log attack time (*LT*), temporal centroid (*TC*) mentioned in timbre model are selected to construct feature set for HS. *SC* reflects the signal power distribution in frequency domain. *LT*, which is also called rising time in other literatures, is of great use in feature extraction on morphology. *TC* reflects the strength of the

Manuscript received March 9, 2014; revised May 9, 2014. This work was supported in part by the Research Committee of University of Macau under Grant MYRG184 (Y1-L3)-FST11-DMC, and Science and Technology Development Fund (FDCT) of Macau under Grant 018/2009/A1, respectively.

Hai-Yang Wang, Guang-Pei Li, Bin-Bin Fu, Jun Huang, and Ming-Chui Dong are with the Electrical and Computer Department, Faculty of Science and Technology, University of Macau, Macau S.A.R., China (e-mail: dylanchriswang@hotmail.com, lee.gp@foxmail.com, ariespleo51@gmail.com, hjwongzeon@gmail.com, mcdong@umac.mo).

DOI: 10.7763/IJBBB.2014.V4.362

signal distribution in time domain. It is the geometric centroid of the signal in temporal domain. The definition and detailed explanation are listed as follows.

SC is computed as the power spectrum weighted average of the frequency in the power spectrum.

$$S(k) = \sqrt{\sum_{i=1}^{M} P_i(k)/M} \tag{1}$$

$$SC = \frac{\sum_{K=1}^{NFFT} f(k) \cdot S(k)}{\sum_{K=1}^{NFFT} S(k)} \tag{2}$$

where M is the total number of frames in a sound segment, $P_i(k)$ is the k^{th} power spectrum coefficient in the i^{th} frame, $f(K)$ is the k^{th} frequency bin.

LT is a very short period of time that elapses before the sound has formed its spectrum. How instruments are identified usually depends on log attack time. It is defined as the logarithm of duration between the time where signal starts and the time where it reaches its stable part.

$$LT = \log_{10}(T_1 - T_0) \tag{3}$$

where the T_0 is the time when the signal starts. Here we use the time while the amplitude is 0.001 db. T_1 is the time the signal reaches its sustained part of maximum part.

TC is defined as time average over the energy envelope of the signal.

$$TC = \frac{\sum_{n=1}^{length(SEnv)} n/s_r \cdot SEnv(n)}{\sum_{n=1}^{length(SEnv)} SEnv(n)} \tag{4}$$

where s_r is the sample rate. The envelope is computed by Hilbert transform (Fig. 1). *TC* reflects the energy distribution in the temporal domain.

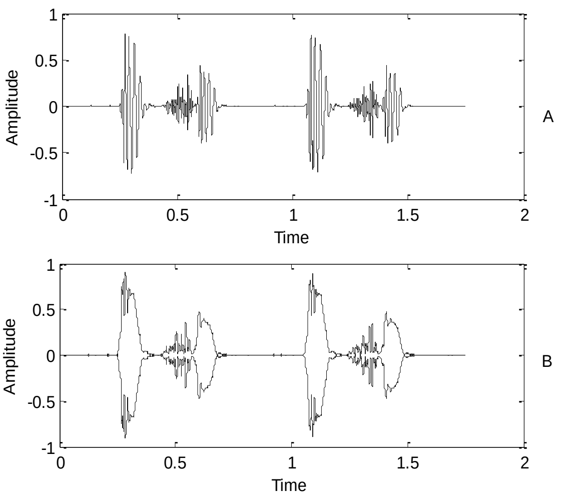

Fig. 1. HS signal and its envelop computed by Hilbert transform; A is the orginal HS signal; B is its envelope computed by Hilbert transform.

III. Test Results

A. Setting of Experiments

To evaluate the validity of defined and extracted multidimensional feature, two experiments are designed. One is to check the effect to characterize the sounds produced by common musical instruments. Here the percussive sounds like drum and cymbal are chosen due to which they are mostly related to the heart sound comparing with sounds produced by other musical instruments. Further the proposed feature extraction is applied to real heart sound dataset including different sound types to see its distinguishable performance.

The data for the first experiment are a set of different seven cymbals and seven drums, comprised of five tom drums and two snare drums. The recorder samples have different loudness levels. Moreover, the signal recorded only reflects one stroke on percussive instruments. The second experiment is based on the HS benchmark database provided by eGeneral Medical Inc., USA [17]. The sound types include ejection, rumble and gallop. Furthermore, one normal HS is used for comparison.

The sounds are digitized using a sampling frequency of 44100 Hz. The spectrograms are computed with Fast Fourier Transform (FFT). All processing of the signals are performed in MATLAB (The Math Works, Inc., Natick, MA, USA).

B. Results and Discussions

1) Experiment on drum and cymbal sounds

TABLE I: Multidimensional Feature Extracted from Drum and Cymbal

		Dimensions		
		TC **(s)**	***LT*** **(s)**	***SC*** **(×10⁴ Hz)**
Drum	1	0.2175	2.6352	0.237
	2	0.0760	0.8132	0.191
	3	0.1288	1.8316	0.261
	4	0.1526	0.6554	0.821
	5	0.4057	1.6066	0.207
	6	0.3632	2.2642	0.180
	7	0.2943	2.5789	0.388
	Mean	0.2340	1.7693	0.326
Cymbal	1	0.2888	2.6053	2.11
	2	0.2918	2.8686	1.88
	3	0.2025	3.2462	2.50
	4	0.8863	3.8400	3.47
	5	1.7590	3.9668	2.62
	6	0.2135	3.1274	3.12
	7	0.0918	3.1337	1.76
	Mean	0.49223	3.2292	2.46

In Table I, the values of multidimensional feature of drum and cymbal are listed. Judged by the data, the results in *SC* dimension show obvious divergence that the values of drum are all around 10^3 Hz while the values of cymbal are all one order bigger than the drum's. As to *LT* dimension, the value range of drum is from 2.6053s to 3.9668s while the results of cymbal are from 0.6554s to 2.6352s, which are different and less overlapped. Moreover, it has been found that the smaller *LT* value is, the stronger the sense of stroke a listener heard. For instance, No.4 and No.5 of cymbal in Table I, both of them sound like a slight rub on the cymbal surfaces by a stick rather than a stroke on them by a stick. Therefore *LT* could be utilized in distinguish HSs among ejection, gallop, rumble etc. by supposing that their dashing degrees are various in acoustic perceptions. With regard to *TC* dimension, the ranges of these two group *TC* values have large intersections according to Table I. *TC* is not ideal in separating cymbal and drum comparatively with aforementioned two dimensions. Taken together, the capability of proposed feature extraction to distinguish drum and cymbal is proven as shown in 3-D effect (Fig. 2).

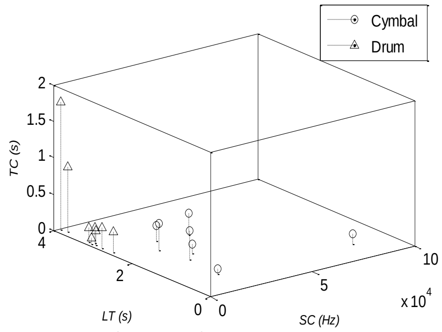

Fig. 2. Multidimensional features of drum and cymbal represented in 3-D diagram.

Although *TC* is not effective in extracting timbre feature of cymbal and drum, it is inherent in interpreting HSs since it takes signal's energy distribution in the temporal domain into account. As shown in Fig. 3, from normal to different abnormal HSs they have obviously different geometric centers, which reflect different energy distribution over the time. Consequently, *TC* provides significant information in finding different aspects of the main pathological HS categories listed in Fig. 3 theoretically. In next section, the results by applying proposed feature extraction to real HS dataset will be presented and discussed.

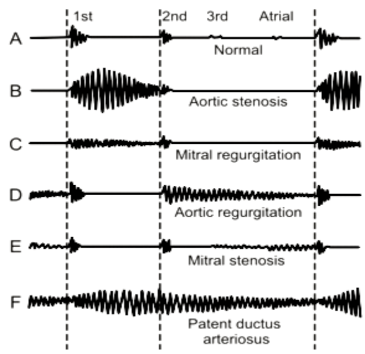

Fig. 3. Normal HS and pathological HS categories.

2) *Experiment on HS dataset*

TABLE II: MULTIDIMENSIONAL FEATURE EXTRACTED FROM 4 TYPES OF HSS

Name	Dimensions		
	***TC*(s)**	***LT*(s)**	***SC*(Hz)**
Ejection	0.2172	2.6245	195.9161
Rumble	0.2582	3.7913	44.9757
Gallop	0.1952	2.7852	121.5734
Normal	0.1472	3.3820	21.3995

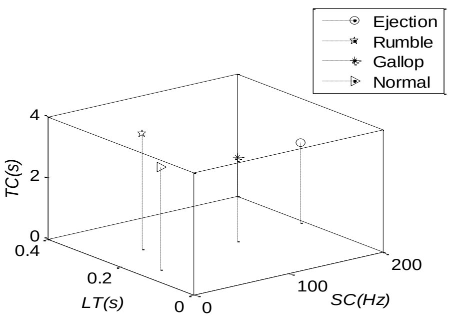

Fig. 4. Multidimensional features of 4 types of HS represented in 3-D diagram.

In this paper, 4 types of HS noted by acoustic descriptive terms are selected. The extracted multidimensional feature values are listed in Table II. As mentioned before; the value of *LT* can be utilized as a reverse indicator for dashing intensity in acoustic perception. In the light of clinical diagnostic basis, the dashing sense of ejection, gallop, normal, and rumble HS is becoming weaker successively, which is proved by the *LT* data shown in Table II. Furthermore, the results in *TC* dimension differ obviously, which is consistent with the assumption discussed in previous section. As a whole, the multidimensional features of 4 types of HS illustrated in 3-D space can be observed with significant differences. They can be applied to those classifiers using special distance as the basis of classification.

IV. CONCLUSION

A multidimensional feature extraction technique is creatively proposed to interpret HS through utilizing three parameters of timbre model: *SC*, *LT* and *TC*.

The experimental results based on musical instruments and real HS dataset are both encouraging. Furthermore, originated from clinical auscultation knowledge, the proposed technique shows better performance and robustness than prevalent segmentation-based HS signal feature extraction methods. Future work can be concluded as exploring the classification method grounded on the extracted multidimensional features.

REFERENCES

[1] Z. Dokur and T. Ölmez, "Dimension reduction by a novel unified scheme using divergence analysis and genetic search," *Digital Signal Processing*, vol. 19, pp. 521-531, 2009.

[2] H. Liang and I. Hartimo, "Feature extraction algorithm based on wavelet packet decomposition for heart sound signals," presented at the IEEE-SP International Symposium 1998.

[3] P. Vikhe, S. Hamde, and N. Nehe, "Wavelet Transform Based Abnormality Analysis of Heart Sound," in *Proc. International Conf. Advances in Computing, Control, & Telecommunication Technologies*, 2009, pp. 367-371.

[4] W. Kao, C. Wei, J. Liu, and P. Hsiao, "Automatic heart sound analysis with short-time Fourier transforms and support vector machines," in *Proc. MWSCAS 9th IEEE International Midwest Symp. Circuits and Systems*, Aug. 2-5, 2009, pp. 188-191.

[5] A. Djebbari and F. Bereksi, "Short-time Fourier transform analysis of the phonocardiogram signal," in *Proc. 7th IEEE International Conf. Electronics, Circuits and Systems*, 2000, pp. 844-847.

[6] H. M Hadi, M. Y. Mashor, M. Z. Suboh, and M. S. Mohamed, "Classification of heart sound based on s-transform and neural network," presented at 10th International Conference, May 10-13, 2010.

[7] M. Varanini, G. Paolis *et al*., "Spectral analysis of cardiovascular tme series by the s-transform," *Computers in Cardiology*, 1997, pp. 383-386.

[8] C. Sunita *et al*., "A computer-aided MFCC-based HMM system for automatic auscultation," *Computers in Biology and Medicine*, vol. 38, pp. 221-233, Feb. 2008.

[9] F. Wenjie, X. Yang, and Y. Wang, "Heart sound diagnosis based on DTW and MFCC," in *Proc. 3rd International Congress*, vol. 6, pp. 2920-2923, 2010.

[10] I. Kamarulafiza *et al.*, "Heart sound analysis using MFCC and time frequency distribution." *World Congress on Medical Physics and Biomedical Engineering 2006*, Springer Berlin Heidelberg, 2007.

[11] J. Grey, "An exploration of musical timbre," Ph.D. dissertation, Dept. of Music Report No. STAN-M-2, Stanford University, CA, 1975.

[12] S. Lakatos, "A common perceptual space for harmonic and percussive timbres," *Perception & Psychophysics*, vol. 62, pp. 1426-1439, 2000.

[13] S. McAdams, "Perspectives on the contribution of timbre to music structure," *Computer Music Journal*, vol. 23, pp. 85-102, 1999.

[14] H. V. G. Evangelista, "Separation of harmonic instruments with overlapping partialsin muti-channel mixtures," presented 2003 IEEE Workshop an Applications of Signal Prwessinp to Audio and Acoustics.
[15] K. Jensen, "Timbre Models of Musical Sounds," Nr. 99/7 ISSN 0107-8283.
[16] A.Caclin, S. McAdams *et al.*, "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," *Acoustical of Society Amecia*, vol. 118, pp. 471, April 2005.
[17] Multidimensional features distinguish them undoubtedly. [Online]. Available: http://www.egeneralmedical.com/listohearmur.html

Hai Yang Wang received the B.Sc degree in automation from Beijing Union University in 2012. He is pursuing a master's degree in the University of Macau. His research interest is biomedical processing and applications etc. His postal address is Faculty of Science and Technology, University of Macau, Macau.

Guang Pei Li is now pursuing for his bachlor's degree of electrical and computer engineering in University of Macau. His interest is in biomedical engineering and microprocessors. His postal address is Faculty of Science and Technology, University of Macau.

Bin Bin Fu is currently pursuing PhD degree in ECE Department of University of Macau. Her research interests focus on embedded system, biomedical signal processing and artificial intelligence, towards the goal of synthesizing them for wearable/portable monitoring with computerized risk stratification.

Jun Huang is currently pursuing master degree in ECE Department of University of Macau. His research interests focus on biomedical signal acquiring and processing with the goal of constructing a comprehensive bio-database.

Ming Chui Dong is a full professor and PhD supervisor of ECE Department of University of Macau, Automation Department of Tsinghua University, China and a professor of YanTai University, China. He received MSc degree in EEE in 1975 at Tsinghua University, China, Visiting Scholar in EEE in 1981 at Rome University, Italy.