Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A Procedure to Extract the Aortic and the Pulmonary Sounds from the Phonocardiogram

Vivek Nigam and Roland Priemer

Electrical and Computer Engineering Department

University of Illinois at Chicago, Chicago, Illinois, USA

Abstract— **The time interval between the aortic (A2) and pulmonary (P2) components of the second heart sound (S2) is an indicator of the presence and severity of several cardiac abnormalities. However, in many cases identification of the A2 and P2 components is difficult due to their temporal overlap and significant spectral similarity. In this work, we present a method to extract the A2 and P2 components from the S2 sound, by assuming their mutual statistical independence. Once extracted, the A2 and P2 components are identified by using a physiological reference signal. Results obtained from real data are encouraging, and show promise for utilizing the proposed method in a clinical setting to non-invasively tract the A2-P2 time interval.**

I. Introduction

The human heart produces an almost periodic 'lub-dub' sound that can be heard with a stethoscope. A recording of this sound is called a phonocardiogram (PCG), which provides an electrical tracing of acoustic energy produced by the heart. The first heart sound (S1), which is the 'lub' of 'lub-dub', is produced primarily by the closure of the mitral and tricuspid valves to prevent backward flow of blood from the ventricles into the auricles during ventricular systole when the ventricles are pumping blood into the aorta and the pulmonary artery. The second heart sound (S2) or the 'dub' is produced by the closure of the aortic and pulmonary valves to prevent the backward flow of blood from major arteries into the ventricles after they have finished pumping blood and enter a state of relaxation. Since right ventricular ejection begins prior to left ventricular ejection and terminates after left ventricular ejection, it has a longer duration, and the sound produced by pulmonary valve closure (P2) normally occurs after the sound produced by aortic valve closure (A2). The time interval between the A2 and P2 components, henceforth referred as tap, is an indicator of the presence and severity of various cardiac abnormalities that include complete right bundle branch block, atrial septal defect, ventricular septal defect, pulmonary stenosis, surgical banding of the pulmonary artery and pulmonary hypertension to name a few [1]. Paradoxical splitting of the A2 and P2 components, in which the A2 sound is delayed and occurs after the P2 sound, is known to occur in left bundle-branch block, aortic stenosis, patent ductus arteriosis and systemic vascular hypertension [1]. In all the above mentioned abnormalities, tap could vary from being as small as 0.01 seconds to being as large as 0.1 seconds or more [1]. Although, visual identification of the A2 and P2 components from the PCG is possible in abnormalities where tap is large, identification of the A2 and P2 components when tap is less than 0.03 seconds is difficult, because of the considerable overlap between them [2].

Xu et al. have developed a method for extracting the A2 and P2 components of the S2 sound, to estimate tap, by using nonlinear transient chirp modeling of the A2 and P2 components [2]. First, the instantaneous frequency of the A2 component is estimated by masking the Wigner-Ville distribution (WVD) of the S2 sound to obtain the phase function and the analytical form of the A2 sound followed by dechirping to obtain the amplitude of the A2 sound. Then, the synthesized A2 component is subtracted from the S2 sound to obtain the P2 sound. However, this method will fail to yield the A2 and P2 components when their instantaneous frequencies overlap significantly in WVD and S2 appears to be a monocomponent sound. Also, the accuracy of reconstruction of the A2 component depends upon the visual masking of the WVD of the S2 sound, to identify energy ridges, and this is prone to human errors in the presence of appreciable cross-energy terms.

This paper presents a method to extract the A2 and P2 components from the S2 sound by assuming that the A2 sound, which is due to the closure of the aortic valve, and the P2 sound, which is due to the closure of the pulmonary valve, are statistically independent. This assumption seems valid because the aorta and the pulmonary artery are two independent arteries in the human circulatory system. Towards the end of ventricular systole when the ventricular pressure diminishes, the backward flow of blood in the aortic root, which causes the closing of the aortic valve to produce the A2 sound, is not conditioned upon the backward flow of blood in the root of pulmonary artery, which causes the closing of the pulmonary valve to produce the P2 sound [3]. This allows us to apply blind source separation (BSS) based algorithms to simultaneously recorded S2 sounds to extract the A2 and P2 components solely on the basis of their mutual independence. Utilizing a suitable physiological reference that is described later, the extracted A2 and P2 components are identified, and then tap can be estimated.

1-4244-0033-3/06/$20.00 ©2006 IEEE.

II. METHODS

A. Blind Source Separation based Extraction of the A2 and P2 Components

BSS denotes observing mixtures of independent sources, and by making use of these mixtures only and nothing else, recover the original signals [4]-[6]. The basic BSS problem assumes instantaneous and noiseless mixing of the sources, and it is modeled by a linear relation between the observations **x** and the sources **s** given by, **x** = **As**, where $\mathbf{x} \in R^n$, $\mathbf{s} \in R^m$, $\mathbf{A} \in R^{n \times m}$. We assume that the components of the source vector **s** are statistically independent and have probability distributions that are not Gaussian except for at most one component [7]. To obtain a unique separation of sources given a set of mixtures, we assume that *m*, the number of sources, is less than or equal to *n*, the number of observations. The goal of BSS is to estimate a separation matrix **W** that satisfies **WA** = **PD,** where $\mathbf{W} \in R^{m \times n}$, $\mathbf{P} \in R^{m \times m}$, $\mathbf{D} \in R^{m \times m}$, and where **P** is a permutation matrix that has only one large entry in each of its rows and columns and **D** is a diagonal matrix. With the separation matrix, we can reconstruct the original source estimates given by **y** = **Wx**. Learning **W** by observing **x** only requires making use of higher order statistics. Typically, the separating matrix **W** is calculated iteratively by optimizing some cost function of the source estimates **y** [7]. In the current work, **s** consists of the A2 and P2 components. The mixing matrix **A** is the unknown transfer matrix that causes the mixing of the A2 and the P2 components to yield the S2 sound. The values of the elements of **A** depend upon the transmission characteristics of human thorax. The observation **x** contains the simultaneous recordings of the S2 sound from acoustic sensors. Therefore, our aim is to estimate the values of **s** given **x** only.

Let x_i denote the output of the i^{th} sensor recording the PCG. In practice, the acoustic recordings x_1 and x_2 will be contaminated with sensor and other physiological noise, which act as additional interfering unknown sources. Hence, we must use more than two simultaneous recordings of the PCG to extract the A2 and P2 components from the S2 sounds isolated from multiple PCG recordings. Let us assume that we have *N* simultaneous recordings of the PCG, $x_1, x_2, \ldots, x_N$. Let t_i_begin and t_i_end denote the start and end of the S2 sound, visually identified, from the i^{th} sensor recording x_i. We define two new quantities as *t_begin* = min $\{t_i\_begin\}$, $i = 1,2,\ldots,N$ and *t_end* = max $\{t_i\_end\}$, $i = 1,2,\ldots,N$. Subsequently, S2 from x_i, which is denoted by $S2_i$ is extracted as $S2_i = x_i$ (*t_begin: t_end*). Once we have isolated the $S2_1, S2_2, \ldots, S2_N$, we form a matrix, $\mathbf{X} \in R^{N \times L}$, as $\mathbf{X} = [S2_1; S2_2; \ldots; S2_N]$, where $L = length(S2_i)$. Finally, the FastICA [7] algorithm is applied on **X** to estimate the independent components of S2, i.e., A2 and P2.

The FastICA algorithm is a batch algorithm that estimates the unknown independent sources by optimizing (maximizing or minimizing) the kurtosis of the source estimates. As a preprocessing step, the observed data is first centered and then whitened by applying a linear transformation that makes the components of **X** uncorrelated with unity variance. Whitening is done by principal component analysis projection: **V** = **ZX**, with $\mathbf{V}\mathbf{V}^T = \mathbf{I}$, where **I** is the identity matrix. The whitening matrix **Z** is given by $\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T$, where $\mathbf{\Lambda} = diag[\lambda(1),\ldots,\lambda(N)]$ is a diagonal matrix with the eigenvalues of the data covariance matrix, $\mathbf{C} = (1/L)\mathbf{X}\mathbf{X}^T$, and **U** is a matrix with the corresponding eigenvectors of **C** as its columns. Whitening transforms the problem of estimating **W** into a simpler problem of finding an orthogonal matrix **B** such that **B** = **ZA**. The FastICA algorithm is a fixed point algorithm that operates on **V** to find the columns of **B** (denoted by **b**), one at a time, and so identifies one independent source at a time. The l^{th} iteration of this algorithm is defined as: $\mathbf{b}_l^* = \mathbf{V}(\mathbf{V}^T\mathbf{b}_{l-1})^3 / L - 3\mathbf{b}_{l-1}$, $\mathbf{b}_l = \mathbf{b}_l^* / \|\mathbf{b}_l^*\|$ [7]. To estimate more than one independent solution, the algorithm can be executed repeatedly by removing from the newly found **b** the projections of previously found columns of **B** [7]. The estimates of the original unknown sources are finally obtained with $\mathbf{Y} = \mathbf{B}^T\mathbf{V}$, where each row of **Y** contains a source estimate.

B. Identification of the A2 and P2 Heart Sounds

The outputs of the FastICA algorithm are arbitrarily permuted and scaled estimates of the original unknown sources. Although the scaling of the source estimates has no affect on the estimation of *tap*, their permutation can make identifying the A2 and P2 components from the output of the ICA algorithm difficult. To alleviate this problem, we record the carotid pulse from the neck using an additional sensor. Because of its proximity to the heart and great vessels, the carotid pulse closely resembles the aortic pressure-pulse and exhibits a dicrotic notch that is produced by the abrupt closure of the aortic valve [1]. The A2 sound ends before the dicrotic notch of the carotid pulse and this relationship tends to remain constant for the same individual [1]. This information is used to identify the A2 sound and to distinguish it from the P2 component in the ICA outputs.

III. RESULTS

We now apply the method described in the last section to extract the A2 and P2 components from the S2 sound. Fig. 1 shows a subject with a harness that applies several stethoscopes to the chest. The stethoscopes were positioned to obtain the PCG from pulmonary, mitral and tricuspid auscultatory locations [1]. Four simultaneous recordings of the heart sound were made by using Littmann® Model 4000 electronic stethoscopes, the outputs of which were subsequently sampled at a 5 KHz rate using the National Instruments model PCI-6251 sixteen channel and 16-bits per sample A/D converter. Four simultaneous recordings of the S2 sound to estimate the A2 and P2 components were used, because attempts to isolate them from two and three simultaneous recordings of the S2 sound failed. We attribute this failure to the presence of more than two independent sources (as assumed), such as extraneous noise and other physiological sounds in the recorded heart

sound mixtures. An Inexpensive Piezoelectric Acoustic Sensor, developed in the Acoustics and Vibrations Lab at UIC, was positioned on the neck to digitize the carotid pulse emanating from the carotid artery.

Subsequently, the PCG recordings were filtered by applying an FIR linear phase shift band-pass filter of length 300 taps, with a passband from 30 Hz -150 Hz, to remove out of band interference due to breathing and other extraneous noise sources. The carotid pulse was also filtered with an FIR low-pass filter with a cutoff frequency of 100 Hz. Then, the S2 sound was isolated from all four recordings using the method described in Section II. Subsequently, the FastICA algorithm was applied to the S2 sounds that were isolated from the simultaneous recordings, to extract the A2 and P2 components. Fig. 2 shows the S2 sounds isolated from four simultaneous recordings of the heart sound, where we see that it is difficult to discern the A2 and P2 components within any one of them. Fig.3 shows the four independent components (ICs) estimated by applying the FastICA algorithm to the four mixtures shown in Fig. 2. Also shown in Fig. 3 is the section of the carotid pulse that lies within the duration of the segmented S2 sounds in Fig. 2. We have manually annotated the dicrotic notch on the carotid pulse. On careful observation we notice that the third IC is very similar to the fourth S2 sound in Fig. 2, and the fourth IC is a mixture of the first and the second ICs and a slowly varying waveform. We deduce that the third and the fourth ICs are extraneous ICs that contain contributions from several physiological sources. This leaves the first and second ICs as the possible candidates for the A2 and P2 components. Notice, how the third IC ends before the occurrence of the dicrotic notch, making it the most likely A2 sound estimate. Similarly the second IC ends after the dicrotic notch, making it the most likely P2 sound estimate.

To estimate *tap*, four sets of A2 and P2 components were extracted from four successive cardiac cycles. Then, the extracted A2 components were time aligned by observing their peaks and then averaged to estimate A2. The extracted P2 components were time aligned following A2 and averaged to estimate P2. These estimates are shown in Fig. 4. Averaging diminishes extraneous noise, resulting in a better estimate of *tap*. Finally, *tap* can be estimated by finding: 1) the difference between the time centroids of the estimates of A2 and P2, 2) the lag value at which the correlation between the estimates of A2 and P2 has its maximum value, and 3) the time between the maximum peak of the estimates of A2 and P2. For the time duration of the estimates of A2 and P2 we use the common time range resulting from aligning A2 and P2. Within this duration, the time centroid of the K samples of A2, for example, was found with: $\sum_{i=1}^{K} (iT)\ abs(\mathrm{A2}(iT)) / \sum_{i=1}^{K} abs(\mathrm{A2}(iT))$, where T is the sampling time interval. Table I gives the values of *tap* estimated by the three methods mentioned above. We notice that while the cross-correlation and maximum peak methods give similar results, the time centroid method yields a larger estimate of tap because it uses the same duration for A2 and P2, which tends to increase the time centroid of P2. The estimate of *tap* will improve as we average over a larger number of extracted A2 and P2. This, however, is limited, because the subject is required, like in most cardiac auscultation, to hold the breath during PCG recording.

IV. Discussion and Conclusion

We have presented a BSS based method to separate the A2 and P2 components of S2 to estimate the time interval between their occurrences. To our knowledge this is the first attempt to isolate heart sound components under the BSS framework. Mutual independence of A2 and P2 was assumed to extract them from S2. The performance of the algorithm does not depend upon the time interval between them, as is the case with the method presented by Xu et al. [2]. We have demonstrated that the proposed method is useful for non-invasively estimating *tap* in a clinical setting. However, no algorithm is without its limitations. The method presented in this paper requires visual identification of the dicrotic notch from the carotid pulse, which might not be easy to do when the notch appears as an inflection point in the presence of noise. Oppenheim and Sittig [8] discuss several algorithms that automate detection of the dicrotic notch in a variety of conditions. The harness shown in Section III could prove cumbersome in clinical applications, and it can be replaced by a stethoscope having several miniature acoustic sensors instead of a single diaphragm. Then, a dedicated microprocessor can be used to process the recorded heart sounds to extract the A2 and P2 components and estimate the time split between them. The isolated A2 waveform can also be utilized to estimate the time of aortic valve closure, which is widely used to determine the duration of systole and the condition of the left ventricle.

Acknowledgement

- Vivek Nigam is supported by a fellowship from the Prakash Agarwal Foundation.
- We thank Todd Spohnholtz and his advisor Dr. T. J. Royston of the Acoustics and Vibrations Laboratory, Univ. of Illinois at Chicago, for providing the Inexpensive Piezoelectric Acoustic Sensors (IPAS) to record the carotid pulse in Section III.

References

[1] M. E. Tavel, *Clinical Phonocardiography and External Pulse Recording,* 4th ed. Chicago, Ill: Yearbook Medical Publishers, 1985.

[2] J. Xu, L.G. Durand and P. Pibarot, "Nonlinear transient chirp signal modeling of the aortic and pulmonary components of the second heart sound," *IEEE Trans Biomed. Eng*, vol. 47, no.10 pp.1328-35, Oct., 2000.

[3] R. F. Rushmer, *Cardiovascular Dynamics*, 4th ed. London: Saunders, 1978, pp. 425

[4] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287- 314, April, 1994.

[5] A. T. Bell and T. J. Sejnowski, "An Information maximization approach to blind separation and blind deconvolution," *Neural Computation,* vol. 7, no. 6, pp. 1004-1034, 1995.

[6] S. Amari, A. Cichocki and H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information processing Systems*, Cambridge, MA: MIT Press, vol. 8, pp. 752-763, 1996.

[7] A. Hyvarinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.

[8] M.I. Oppenheim and D.F. Sittig, "An innovative dicrotic notch detection algorithm which combines rule-based logic with digital signal processing techniques," *Computers and Biomedical Research,* vol. 28, Issue 2, pp.154-170, April 1995.

TABLE I: VALUE OF TAP FOR A2 AND P2 COMPONENTS IN FIG. 4

Method for tap estimation	**Estimated *tap***
Centroid based	19 msec
Cross-correlation based	15.6 msec
Time between maximum of A2 and P2 sounds	15.8 msec

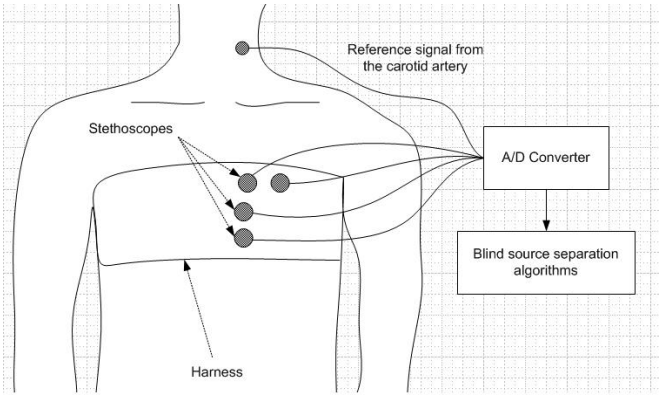

Fig. 1: The experimental setup to simultaneously record the PCGs and the carotid pulse.

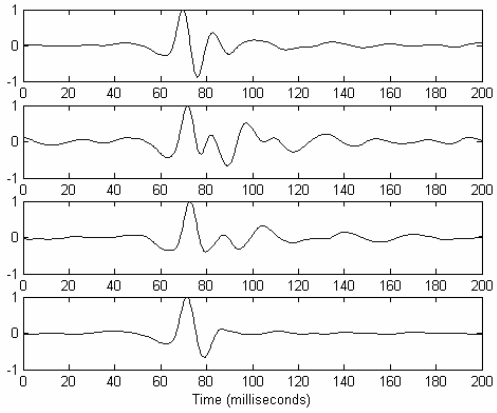

Fig. 2: The four S2 sounds recorded with the apparatus shown in Fig. 1.

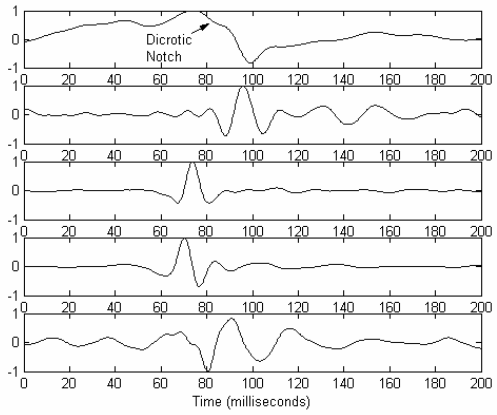

Fig. 3: The carotid pulse (top) and the four ICs extracted from the four mixtures shown in Fig. 2.

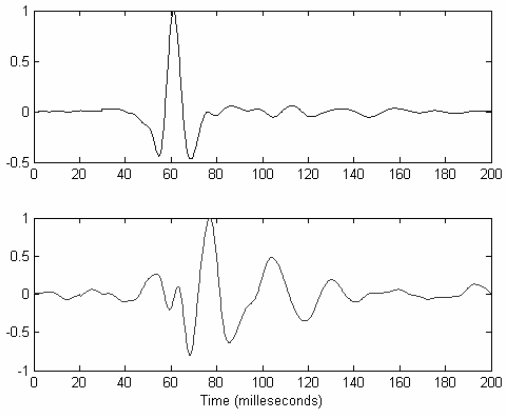

Fig. 4: The estimated A2 component (top) and the estimated P2 component (bottom).