International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 6

ISSN: 2321-8169
4257 – 4260

Decorrelation of Lung and Heart Sound

A. A. Ballewar
Department of Electronics & Telecom.
Sinhgad College of Engineering
Pune, India
anurag.ballewar@gmail.com

Dr. S. D. Lokhande
Department of Electronics & Telecomm.
Sinhgad College of Engineering
Pune, India
sdlokhande.scoe@sinhgad.edu

Dr. R. R. Kodgule
Head Academic Research
Chest Research Foundation
Pune (MH), India
rahulkodgule@gmail.com

***Abstract*— Signal separation is very useful where several signals have been mixed together to form combined signal and our objective is to recover individual original component signals from that combined signal. One of the major problem in neural network and research in other disciplines is finding a suitable representation of multivariate data, i.e. random vectors. For concept and computational simplicity representation is in terms of linear transformation of the original data. This means that each component of the representation is a linear combination of the original variables. There are linear transformation methods such as principal component analysis and Independent Component Analysis (ICA). ICA is a recently developed method in which the goal is to find a linear representation of non-gaussian data so that the components are statistically independent or as independent as possible.

Keywords- *Independent Component Analysis, Blind Signal Separation, Mixing Matrix, Lung Sound Signal, Heart Sound Signal, Decorrelation and Blind Source Separation*

*****

I. INTRODUCTION

Most of the signals from man-made systems such as physical and physiological systems, mechanical, radar and sonar systems, some parameters and statistical properties vary with time. Similarly sound observed from outer layer of chest cavity consists of mixed sound signal and varies according to physical condition of patients. This mixed sound signal consists of signal from lung and heart. Auscultation is the process used by physician which requires lot of experience for examining heart and lung functioning. For this physicians are required to place stethoscope at correct place on chest envelope. Digital stethoscope can also be used to capture variation in this mixed signal. This signal is used for its computer analysis and separation. Here main task is to successfully capture the mixed sound signal and separate them for their proper analysis.

Lung sound while recording breadth sound consists of quasi-periodic heart sound [1][2]. They have overlapping frequency spectrum in the range of 20-150 Hz and in this range lung sound have major components. Therefore cancelation of one of either sound or separation of this mixed sound into their individual source is really a difficult task. As the Heart Sound and Lung Sound overlap in frequency, the required information may be corrupted leading to misinterpretation resulting in wrong diagnosis by the cardiologist.

The cardiac cycle consists of two phases for four chambers of the heart [3] which is the contraction and relaxation also known as systole and diastole respectively. During contraction chamber pushes blood into an adjacent chamber where as in relaxation the chamber relaxes and gets filled with blood. The P wave is related with blood being pushed by atrial contraction into lower two ventricular chambers. The waves Q, R and S form QRS complex and is related with contraction of the ventricles due to ventricular depolarization. The T wave corresponds to repolarization of the ventricles, which allows the ventricles to relax prior to the next cardiac cycle. The first heart sound S_1 occurs immediately after R wave. It is produced as a result of ventricular contraction causing blood to flow back towards the atria. The second heart sound S_2 can be heard at the end of the T wave which is produced by the relaxation of the ventricles causing blood to flow back into these chambers from the arteries.

Lung sound is produced during inspiration and expiration cycles in the frequency range of 20-1200Hz. Normal lung sounds originate from bronchi, bronchioles and alveoli of the lung during inspiration and from central airways (trachea) during expiration. Wheeze and Crackles are two types of Abnormal or adventitious sounds in lungs. Wheezes are continuous abnormal lung sounds with frequency less than 100Hz to more than 1000Hz. They are produced because of walls of narrowed airway and is heard in person with airway obstruction. Crackles are classified as high pitched or fine crackles and low pitched crackles. High pitched crackles are heard in patients with pneumonia or during early stages of congestive heart failure, interstitial pulmonary fibrosis and in patients with chronic obstructive lung diseases the low pitched crackles are observed.

Fatma Ayari et al. [1] introduced an ICA technique which can be used to separate heart and lung sound signal. ICA is used for blind source separation which provides separation of signals from mixed signals with very less information about source signals or mixing method. ICA basically makes signals as independent from each other and decorrelates the signal to reduces higher-order statistical dependencies. The ICA algorithm is based on the well known Cocktail-Party-Problem; if we consider the n source signals S_1, S_2 …. S_n as mentioned in fig. 1 as

ISSN: 2321-8169
4257 – 4260

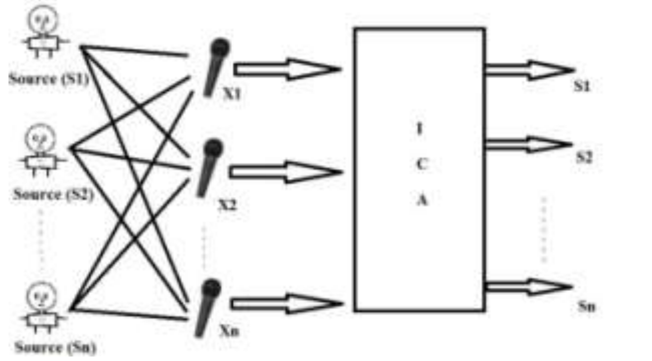


Fig. 1: Independent Component Analysis (ICA) for n-sources [1]

ICA algorithm was developed and simulated in MATLAB software. The Paper is organized as follows- Firstly, Fast-ICA algorithm is described in details in Section II. The system model for separation of mixed sound signals is explained in Section III. In Section IV, results such as Entropy and magnitude of gradient entropy are presented and conclusion is given in section IV.

II. INDEPENDENT COMPONENT ANALYSIS

Consider two microphones placed at different locations. Using microphones we capture two time recorded signals $x_1(t)$ and $x_2(t)$, at time t and amplitudes x_1 and x_2. These recorded signals are weighted sum of the speech signals emitted by the two speakers, which we denote by $s_1(t)$ and $s_2(t)$. We can express it as a linear equation:

$$x_1(t) = a_{11}S_1(t) + a_{12}S_2(t) \qquad (1)$$

$$x_2(t) = a_{21}S_1(t) + a_{22}S_2(t) \qquad (2)$$

where a_{11}, a_{12}, a_{21} and a_{22} are some parameters which depend on the distances of the microphones from speakers. Using only the recorded signals $x_1(t)$ and $x_2(t)$ if two original speech signals $s_1(t)$ and $s_2(t)$ can be estimated then it will be very useful. Generalized form of fig.1 for n independent component can be written as

$$x_j = a_{j1}s_1 + a_{j2}s_2 + a_{j3}s_3 + \cdots + a_{jn}s_n \qquad (3)$$

Here in ICA time index t is dropped; and we assume that each mixture x_j and each independent component s_k is a random variable instead of a proper time signal. Equations (1) and (2) can also be written in more simplified way as-

$$X = A.S \qquad (4)$$

where A is a mixing matrix, X is the recorded signals matrix and S is original sound sources matrix. Here we have mainly two task as mixing matrix A is also unknown, estimate matrixes A and S using only the observable matrix X. For simplicity, we assume that the unknown mixing matrix is square, but this is sometimes relaxed. After estimating the matrix A, we can compute its inverse denoted W and obtain the independent component simply by:

$$S = A^{-1}.X = W.X \qquad (5)$$

By considering two scalar-valued random variables y_1 and y_2 the concept of independence can be easily explained. The variables y_1 and y_2 are assumed to be independent if information on the value of y_1 does not contain any information on the value of y_2 and vice versa, which is similar the variables S_1 and S_2 but not with the mixtures variables x_1 and x_2. Probability density of y_1 and y_2 is written as $p(y_1, y_2)$ and y1 and y2 are independent if and only if the joint probability density function is factorizable as shown below:

$$p(y_1, y_2) = p_1(y_1).p_2(y_2) \qquad (6)$$

this definition extends naturally for any n number of random variables where joint density is product of n terms. Uncorrelatedness is weaker form of independence. If two random variable are independent it means that they are uncorrelated but when two variable are uncorrelated it does not imply independence. The above definition can be used to derive a most important property of independent random variables. Given two functions h_1 and h_2 we have:

$$E\{h_1(y_1)\, h_2(y_2)\} = E\{h_1(y_1)\}\, E\{h_2(y_2)\} \qquad (7)$$

There are two ambiguities in ICA, first is that we cannot determine variances (energies) of each independent component so unit variance is assumed because any scalar multiplier with source s_i always gets canceled when divided by corresponding column a_i. Second ambiguity is we cannot determine the order of independent components, so a permutation matrix P and its inverse is substituted in model to give $x=APP^{-1}s$. The elements of Ps are the original independent variables s_j but in another order and matrix AP^{-1} is just a new unknown mixing matrix to be solved by the ICA algorithms.

For ICA to be possible the fundamental restriction is that the independent components (IC's) must be non-gaussian because Gaussian variables makes ICA impossible. To see why Gaussian are forbidden the joint density of gaussian variable x_1 and x_2 given by equation (8) is shown in figure 2 which are having unit variance and are uncorrelated.

$$P(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \qquad (8)$$

It can be seen that the density is completely symmetric and do not contain any information about directions of the columns of the mixing matrix **A**, so **A** cannot be estimated.

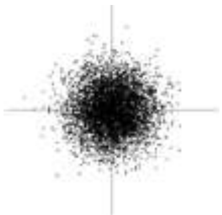


Figure 2: Joint distribution of two independent gaussian variables x_1 and x_2.

This shows that non-gaussianity is key to ICA without which estimation of source is not at all possible. Random variables were assumed to have gaussian distributions by most of classical statistical theory which caused late emergence of ICA. For estimating one of the independent components we consider linear combination of x_i denoted as:

$$y = w^T x = \sum_i w_i x_i \tag{9}$$

$$y = w^T x = w^T As = z^T s \tag{10}$$

Since a sum of even two independent random variables is more gaussian than the original variables, z^Ts is more gaussian than any of the s_i and becomes least gaussian when it in fact equals one of the s_i. In this case, obviously only one of the elements z_i of z is nonzero.

In ICA method first we have to preprocess i.e. centering and whitening of data. Then apply Fast ICA algorithm [1] which can be simplified by using following algorithm:

1. Centering the data to achieve zero mean the result of this step provides zero mean data.
2. Whiten the data which is used to remove the correlation between the observed data. Common method to achieve whitening is Eigen-value decomposition.
3. Choose an initial vector random V of unit norm
4. Calculate : $V_+ = E\{X.g(VT.X)\} - E\{g'(VT.X\}.V$
5. Then normalize it as : $V = V_+ / |V_+|$
6. If not converged, we can go back to step 4.

Convergence means that value of V of current iteration is similar to previous iteration. The function g(.) should have different forms: $g_1(u) = \tanh(a_1 .u)$ where a_1 is any value that is placed between 1 and 2. In step 1 and 2 denoising work is done to attenuate the ambient noise superimposed to the original recorded signals. These preprocessing steps are necessary as they contribute with IC method to improve results.

III. MODEL FOR SEPARATION OF SIGNAL

Let's consider two mixed signals x_1 and x_2 collected via a digital stethoscope by experienced doctors in chest research foundation. Experimental data matrix X is build with signals x_1 and x_2 recorded with a digital stethoscope on a normal patient. Location of digital stethosciope is defined so that the two signals are contain heart and lung signals in such a way that one of them is more affluent than the other with the lung signal and the second is located so that it could be wealthier with heart signal. So the sound signal recording is taken from left and right chest.

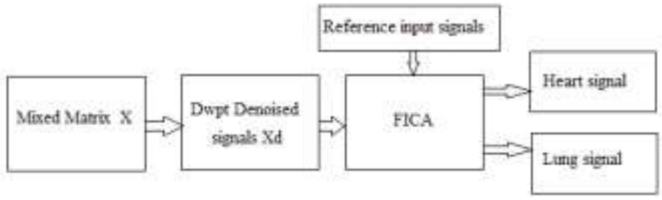
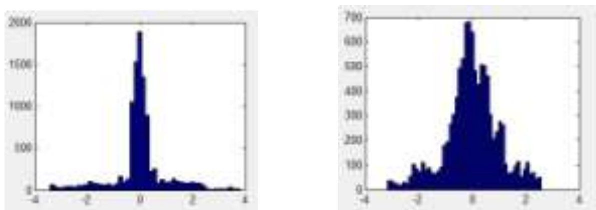


Fig. 3: Scheme of used technique with ICA method [1]

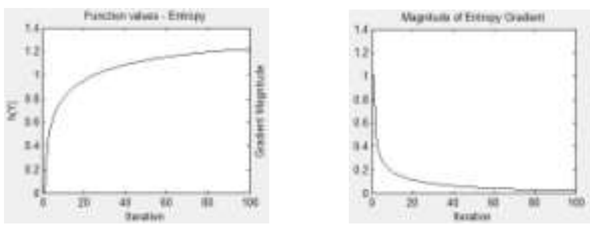
As shown in Figure 3 the recorded signal is used to find matrix X, which is then cleaned from ambient noise using the (dwpt) wavelet packet denoising technique. ICA algorithm is then applied to separate heart and lung sound signals. This Fast ICA algorithm is based on a convolutive sphering process allows use of the classical Fast ICA updates to extract iteratively the innovation processes of the sources in a reduction procedure. It is possible to apply the FICA algorithm more than one cycle time with as far as the outputs of the first cycle is not usually purely separated lung and heart sounds. In [7] and [8] different classification scheme has been proposed to classify crackles based on waveform features and frequency domain features and computer analysis. This purpose is very important in the analysis of respiratory disorders.

IV. RESULTS



(a) (b)

Figure 4: (a) Histogram of source S_1 and (b) Histogram of source S_2



(a) (b)

Figure 5: (a) Entropy of source signal and (b) Magnitude of gradient Entropy.

Figure 4(a) and 4(b) is graphical representation of initial magnitude contained in sound source S_1 and S_2 respectively. The actual distributions of the sound signal can be measured as a histogram of the real and imaginary parts or as histogram of the magnitudes.

The optimization process results are shown in Figure 5(a) and 5(b) where entropy function h(Y) and magnitude of gradient entropy respectively where calculated for 100 iterations. The result in figure 5(a) indicates that as the number of iterations increases the entropy value also increases. This shows that the information content in sound signal in increasing as number of iterations is increasing. After certain number of iteration the increase in number of iteration will not affect the result as graph will become constant, so selection of number of iterations must be sufficient to fulfill our need. The figure 5(b) shows that as number of iterations increases the magnitude of gradient reduces which indicates that variability of source is decreasing and they are converging towards independent sources. This implies that the sources are approaching towards uncorrelation.

V. CONCLUSION

ICA is general-purpose statistical technique in which observed random data is linearly transformed into components which are maximally independent from each other. The results obtained are reduction in correlation which will further lead to separation of individual lung and heart sound. When we apply Fast ICA loop for 100 iterations as see figure 5(a) the entropy of signal increases as number of iterations increases. Magnitude of gradient entropy reduces figure 5(b) which leads in reduction of correlation and convergence to independent sources.

ACKNOWLEDGMENT

I would like to thank my guide for their constant guidance and encouragement in my work. Also I would like to thank Dr.Ashok Aher and Chest Research Foundation, pune for supporting my work.

REFERENCES

[1] Fatma Ayari, Ali T. Alouani, "Lung Sound extraction from mixed lung and heart sound FASTICA algorithm", Electrotechnical Conference (MELECON), 2012 16th IEEE Mediterranean.
[2] Jen-Chien Chien, Ming-Chuan Huang, Yue-Der Lin,Fok-ching Chong, "A Study of Heart Sound and Lung Sound Separation by Independent Component Analysis Technique", Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE.
[3] Vivek Nigam, Roland Priemer, "A procedure to extact the Aortic and Pumonary sounds from the Phonocardiogram", Engineering in Medicine and Biology Society, 2006 EMBS '06. 28th Annual International Conference of the IEEE.
[4] Aapo Hyvarinen, Erkki Oja, "Independent Component Analysis: Algorithm and Applications", Neural Networks, 13(4-5):411-430, 2000.
[5] Sudo T., Tanaka H., Sugimoto C. and Kohno R, "A Study on Single-channel Non-stationary Noise Suppression for Cardiac Sound", Medical Information and Communication Technology (ISMICT), 2014 8th International Symposium.
[6] Sepideh Babaei and Amir Geranmayeh, "Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals", Computers in Biology and Medicine 39 (2009), ELSEVIER.
[7] I. Turkoglua, A. Arslanb and E. Ilkayc, "An expert system for diagnosis of the heart valve diseases", Expert Systems with Applications 23 (2002) 229–236, ELSEVIER.
[8] Pankaj B. Gadge, Bipin D. Mokal, Uttam R. Bagal, "Respiratory Sound Analysis using MATLAB", International Journal of Scientific & Engineering Research Volume 3, Issue 5, May-2012 1 ISSN 2229-5518
[9] Fatma Ayari, Mekki Ksouri, Ali Alouani, "A new scheme for automatic classification of pathologic lung sounds", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012, ISSN (Online): 1694-0814.
[10] G.D. Clifford, "Singular Value Decomposition & Independent Component Analysis for Blind Source Separation", HST582J/6.555J/16.456J Biomedical Signal and Image Processing Spring 2005.
[11] A.Wims Magdalene Mary, Anto Prem Kumar, Anish Abraham Chacko, "BLIND SOURCE SEPARATION USING WAVELETS," IEEE International Conference on Computational Intelligence and Computing Research, 2010.
[12] Hongjuan Zhang, Guinan Wang, Pingmei Cai, Zikai Wu, Shuxue Ding, "A fast blind source separation algorithm based on the temporal structure of signals", Neurocomputing 139 (2014) 261–271, ELSEVIER.