MULTICHANNEL PHONOCARDIOGRAM SOURCE SEPARATION AND LOCALIZATION

Conor Fearon*, Scott Rickard* and Raymond L. Watrous**

* Sparse Signal Processing Group, University College Dublin, Ireland.
** Zargis Medical Corp., Princeton, NJ, USA.

conor.fearon@ee.ucd.ie, scott.rickard@ee.ucd.ie, watrous@zargis.com

Abstract: This work discusses the potential of developing a method which uses multichannel heart sound recordings and sparse representations to localize heart sound energy and separate the phonocardiogram into its constituent components.

Introduction

Auscultation of the heart has long been a powerful yet cost-effective procedure that provides physicians with information on the timing, duration, pitch, location and intensity of normal heart sounds, thus enabling them to make an initial diagnosis or appropriate referral for additional tests. However, advanced technologies have become so much the standard in the practice of cardiology that auscultation has become a neglected art. Mangione et. al. suggests that the resulting low emphasis on cardiac auscultation appears to have affected the proficiency of medical trainees and this has diminished the confidence of many physicians in their ability to accurately detect abnormal cardiac findings through physical examinations [1].

The clinical importance of auscultation cannot be denied, yet accurate detection of pathologies is inherently difficult. As can be seen from Figure 1, a large proportion of heart sound energy is subaudible. Moreover, the sounds are of a low signal amplitude (due to impedance mismatch between the chest and the stethoscope) and obscured by multiple noise sources (breathing, crying, coughing, bowel sounds, speaking, motion artifact). Furthermore, the subjectivity of auscultation leads to inter-observer variability, since individual listeners have different auditory apparatuses. Subjectivity also leads to impressionistic descriptions of murmur characteristics, which makes teaching/learning the skill of auscultation difficult. Sequential listening, coupled with subjective judgments of loudness and pitch and a very limited auditory memory, makes a comparison of different auscultatory sites or different postures challenging, and comparison of sounds from different visits nearly impossible.

Hence, a cost-effective device that would assist the physician in discerning heart sounds and would provide physicians with additional information in support of their decision, would both increase the accurate detection of cardiac abnormalities and reduce the number of unnecessary referrals for more expensive and sometimes invasive procedures.

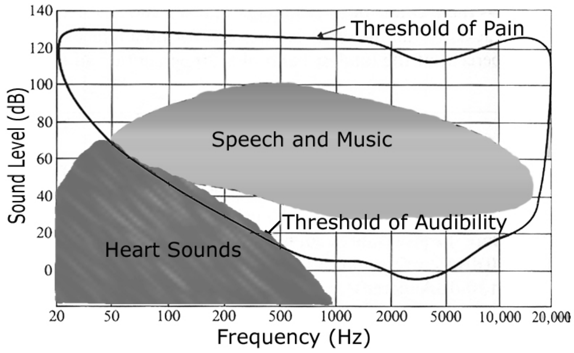

Figure 1: Relationship between the range of sounds produced in the heart and the threshlod of audibility of the human ear.

Currently, the only computer-aided medical device which provides support physicians in identifying heart sounds and murmurs is the Zargis Acoustic Cardioscan [2]. Whereas this system analyzes the timing, intensity and frequency content of heart sounds, it utilizes sequential single channel recordings, in keeping with standard auscultatory protocol, and thus, does not derive location information. This work discusses the potential of developing a method which uses multichannel heart sound recordings to localize heart sound energy and separate the phonocardiogram into its constituent components.

The paper is layed out as follows: the next section briefly describes the mechanisms of heart sound generation. The third section discusses the problem of source separation and describes a number of solutions. In the fourth section, an algorithm for source separation of multichannel heart sounds recordings is introduced. The fifth section presents some preliminary results and the final section contains conclusions and discussions.

Heart Sound Generation

Determining the exact mechanisms implied in the production of heart sounds is a very challenging engineering problem. As a result, substantial controversy still exists concerning the genesis of heart sounds. It is known, however, that the first and second heart sound (S1 and S2) consist of the superposition of two components: M1

ISSN: 1727-1983 © 2005 IFMBE

(left) produced by mitral valve closure and T1 (right) by tricuspid valve closure, in the case of S1; and A2 (left) generated by the closure of the atrial valve and P2 (right) by that of the pulmonic valve, in the case of S2. In certain pathologies other heart sounds may also be audible e.g. systolic ejection sounds which correspond to the opening of the atrioventricular valves; the third heart sound (S3) associated with rapid deceleration during the passive filling phase in early diastole; the fourth heart sound (S4) which is a late diastolic filling sound occuring during atrial contraction; diastolic and systolic murmurs, which are caused by turbulent flow of blood, often taking place at or near a valve but frequently emanating from another part of the cardiac mass and this is to name but a few of the possible heart sounds which can make the phonocardiogram a very complicated signal.

Traditionally, the physician generally listens to the heart sounds by placing a stethoscope at four main auscultatory sites on the chest surface, these are shown in Figure 2. These sites have been chosen because at each one, sounds associated with one of the four valves is heard more clearly. The main problem arising in the separation of heart sounds arises from the fact that a great number of interfering sources are always present. External noise sources include air conditioning, door closings, suction pumps, conversation, alarms etc. More, importantly internal interfering sources are present all around the heart: lung sounds, coughing, bowel sounds, speaking etc. This makes the problem of separating low amplitude heart sound sources considerably more difficult.

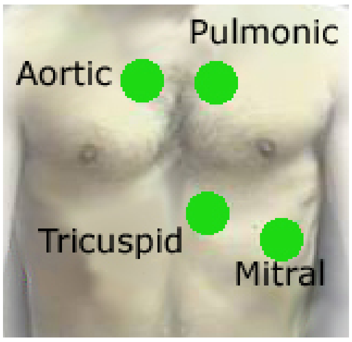

Figure 2: The four main auscultatory sites.

Source Separation Technique

Attempts at localization of heart sounds using multichannel recordings have already been presented in [3, 4]. These methods rely on a subspace method known as MUSIC and require the a priori knowledge of the array response which is difficult in practice given the variability of patient anatomy. A more viable approach involves using histogram-based source separation techniques which rely on sparse representations. This framework has proven to be extremely powerful in the blind source separation of speech [5, 6, 7, 8, 9].

These algorithms involve selecting a representation of the signal which is sparse. Unfortunately there exists a vast number of sparsity measures, which leads to a certain amount of ambiguity when defining sparsity. For all useful purposes, a sparse signal can be be defined as one in which a small percentage of components contains a large perentage of the energy. If this is the case, we can assume that, in a mixture, the likelihood of more than one source having a powerful component at any point in the representation is small. If this is true, then an accurate estimate of the sources can be obtained by using some parameter estimation technique at each point in the representation and applying binary masks to the representation of the mixture.

In more detail, since the probability of sources overlapping in the representation is small but non-zero and noise is present in real signals, the parameters calculated at each point are only estimates of the true values. Accurate estimates of the true values can be obtained by placing all of the parameter estimates in a power-weighted histogram. This leverages the assumption that the points containing most of the energy of each source are disjoint. We expect the results to cluster around the true values of the actual mixing parameters. N sources produces N pairs of mixing parameters which creates N peaks in the histogram. We can then use these true values to partition the representation of a mixture and transform back to the time-domain if necessary.

Given the time-disjoint nature of heart sounds, the time or time-scale domain would give suitable sparse representations (unlike time-frequency, as used in [3]). Unfortunately, this would make it difficult to leverage the phase information introduced by the relative delays of the sources. Assuming a speed of sound in the chest cavity of 1530 m/s, the delays associated with the propagation paths are small (≤ 0.2 msec). Therefore we can assume instantaneous mixing and thus, simpler methods may be employed to separate based on relative amplitude [9].

In the instantaneous source separation problem the mixing matrix **A** consists simply of scalars representing signal amplitudes and the sources will be characterised by a spatial signature consisting simply of amplitudes. So if we have two mixtures x_1, x_2 and three sources s_1, s_2, s_3, the mixing equation would be as follows:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix} \quad (1)$$

If the signals of interest are not sparse in the time domain but are sparse in some other domain and if one creates a scatter plot of x_1 versus x_2 in the time domain, one might get a plot similar to that on the left of Figure 3. However, in the sparse domain, it is unlikely that more than one source is active at any given point, so the linear mixing which results creates lines in the scatter plot (as shown on the right of Figure 3) whose orientations are determined by the columns of the mixing matrix. Noise and occasional simultaneous activity causes the lines to broaden slightly and some outliers to occur, but in the non-sparse domain, the considerable amount of simultaneous activity makes line orientations indistinguishable. A clustering approach could be used to find the line orientations and thus the mixing matrix. Demixing is then possible by partitioning that domain, assigning points to sources based on which line they are closest to [10, 11].

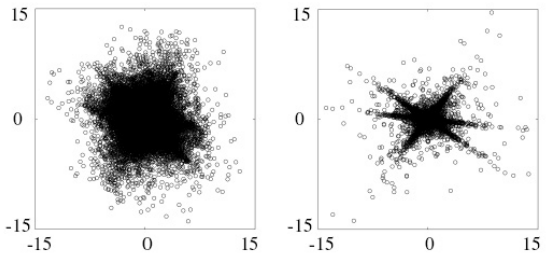

Figure 3: Scatter plot of mixtures in time domain (left) and a sparse domain (right), from [12]

The algorithm which we present in this paper, performs separation in the time-domain and instead of clustering lines in scatter plots, the algorithm calculates the six pairwise ratios of the four auscultatory mixtures and creates power-weighted histogram of the resulting ratios, in the same vein as the DUET Algorithm [5, 6, 7]. Using a power-weighted histogram leverages the assumption that the points containing most of the energy of the sources are disjoint. In the same way that noise and occasional simultaneous activity caused broadening of the lines in the scatter plot, so to does they cause broad peaks in the histogram. A weighted k-Means clustering algorithm is used to locate the centre of the peaks which should be close to the true values. The time domain representation of any mixture can then be segmented based on which peak each time point is closest to.

Results

As an initial test we modelled the valve closure sounds as Daubechies wavelets. As a proof of concept, we introduced attenuations and short but distinct delays between the components of S1 and S2 determined by their location in the thorax and created synthetic mixtures at the four main auscultory sites. We assumed a homogeneous channel with a speed of sound of 1530m/s. Results are displayed below. This model will be extended to include heart murmurs (modelled as coloured noise), amplitude modulation introduced by breathing and noise.

We employed the algorithm for a synthetic mixture with realistic timings between components. The mixtures and resulting demixtures are shown in Figure 1. After segmentation, a suitable angle of arrival estimation can be performed for each mixture pair to determine the exact location of each source.

In real signals, significantly more difficulties arise. Firstly, we have a considerable amount of noise, (motion

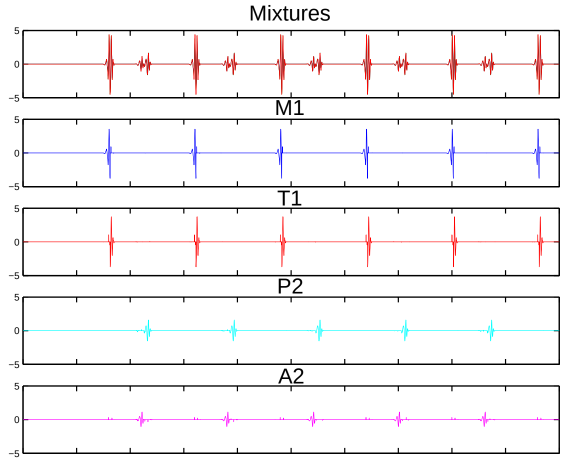

Figure 4: Synthetic mixtures and resulting demixtures.

artefact, external sources) and interfering sources from other parts of the body as outlined above. Furthermore, the instanteous mixing assumption may not be valid. If this is the case (i.e. that we have non-zero relative delays between sensors) then each source will not have constant characteristic relative amplitude, but amplitudes which vary with time as the source signals as heard at all of the sensors fall in and out of phase. Finally we must deal with the fact the sources are not stationary (i.e. their location changes) during the interval of observation, which makes separation based on localization more difficult. As a result of all of these, attempting to separate a real signal is extremely challenging.

To demonstrate the potential for source separation of real signals using this technique, we recorded a 4-channel phonocardiogram using a 4-channel data acquisition system and 4 electronic stethoscopes, from a nominally healthy patient. We then separated this signal by hand into a signal containing just S1 components and another signal containing just S2 components. We then created three two-dimensional power-weighted histograms for each, from the 6 pairwise ratios, as outlined above. One of these is shown for each of S1 and S2 in the upper plots in Figure 5. The disjoint nature of these two components in the time domain is demonstrated by lower plot of Figure 5 which shows the points in the S1 histogram which overlap with those in the S2 histogram. We can assume that the tiny number of points present in the intersection result from some other internal source which is active during both S1 and S2.

The histogram for S2 contains two clear peaks which indicates the presence of two components with different spatial signatures, which would be consistent with the physiology. The histogram for S1 is slightly more difficult to interpret since a number of peaks are visible. S1 will always be more difficult to separate since its components are generally spaced more closely in time but also since these sources are anatomically closer. In both histograms, many of the peaks have tails which is indica-

ISSN: 1727-1983

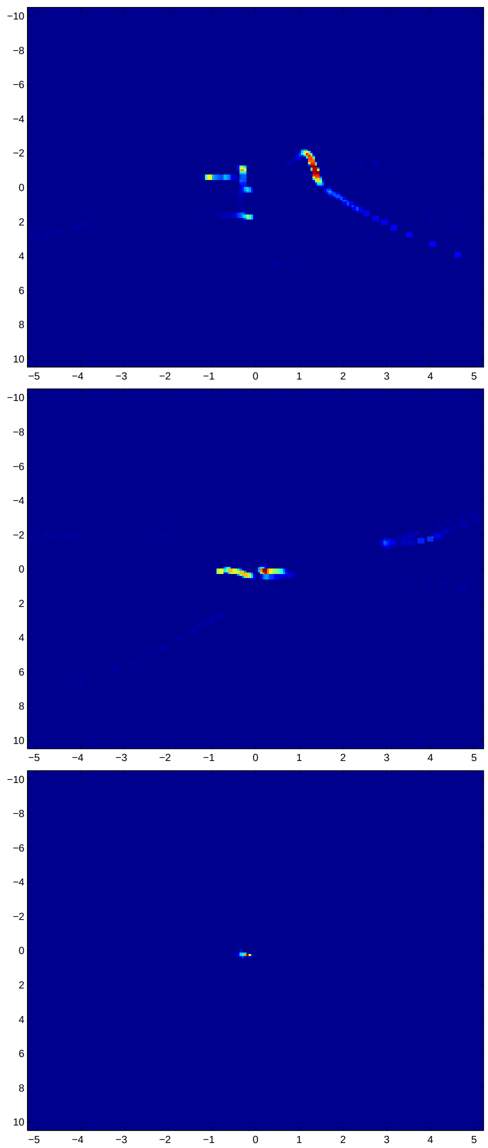

Figure 5: Histograms created from real signal where only S1 is active (top left) and where only S2 is active (top right) and a histogram displaying the points where these two overlap.

tive not of simultaneous activity or noise but possibly movement of the sources during the cardiac cycle or more probably of non-zero delays as discussed above. Further work is necessary to determine the exact cause of these tails. Using the S1 and S2 histograms above, the points which are active during S1 and S2 were determined and thus two masks were created, one for S1 and one for S2. These masks were then applied to one of the mixtures. The estimates for S1 and S2 are shown in Figure 6

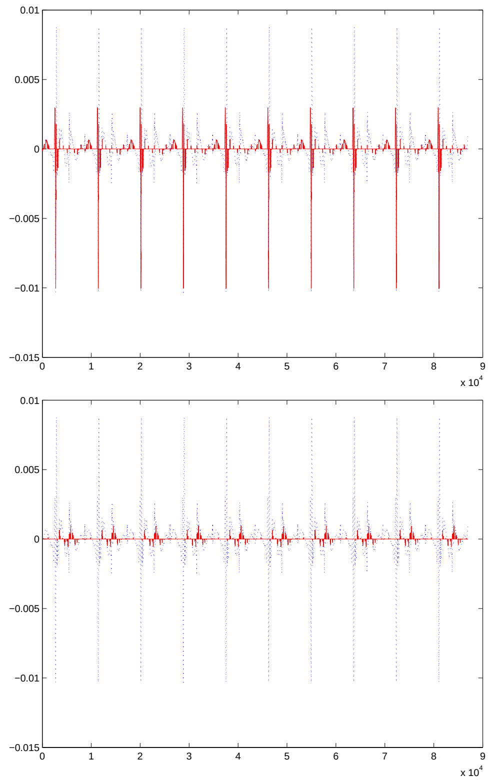

Figure 6: Demixtures for S1 (top) and S2 (bottom) are shown in solid red. The original mixture is also shown in dotted blue.

Conclusions and Discussion

In this work, we discussed the potential of developing a method which uses multichannel heart sound recordings and sparse representations to localize heart sound energy and separate the phonocardiogram into its constituent components. Such technology would have many important applications. The ability to distinguish abnormalities of the mitral or aortic valves from those of the tricuspid and pulmonic valves as well as location of clicks and murmurs would improve the ability of a physician to detect a wide range of pathologies without the need for more expensive and invasive procedures.

The results displayed are only preliminary and further work is required to facilitate separation of sources in real signals. These pilot tests simply demonstrate that once a sparse representation for heart sounds is found separation is possible. Despite the fact that heart sounds are reasonably time-disjoint, considering the overlapping nature of the components of S1, the time-scale domain would seem to be a more effective domain to work in given its noise cancelling properties and redundant representation.

References

[1] S. MANGIONE, L.Z. NIEMAN, E. GRACELY, and D. KAYE. The teaching and practice of cardiac auscultation during internal medicine and cardiology training: A nationwide survey. *Ann. Intern. Med.*, 119(1), 1993.

[2] www.zargis.com.

[3] Y. BAHADIRLAR and H.O. GULCUR. Time-frequency cardiac passive acoustic localization. Proc. 23rd Annual EMBS International Conference, 2001.

[4] Y. BAHADIRLAR and H.O. GULCUR. Cardiac passive acoustic localization: Cardiopal. *ELEKTRIK*, 6(3), 1998.

[5] OZGUR YILMAZ and SCOTT RICKARD. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions Signal Processing*, 52(7), 2004.

[6] S. RICKARD and O. YILMAZ. On the approximate W-disjoint orthogonality of speech. ICASSP2002, Orlando, Florida, USA., 2002.

[7] A. JOURJINE, S. RICKARD, and O. YILMAZ. Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures. *Proc. ICASSP2000*, 5:2985–2988, 2000.

[8] T. MELIA, S. RICKARD, and C. FEARON. Histogram-based blind source separation of more sources than sensors using a duet-esprit technique. Proc. European Signal Processing Conference, 2005.

[9] M. VAN HULLE. Clustering approach to square and non-square blind source separation. IEEE Workshop on Neural Networks for Signal Processing, 1999.

[10] P. O'GRADY and B. PEARLMUTTER. Soft-lost: Em on a mixture of oriented lines. International Conference on Independent Component Analysis, 2004.

[11] P. O'GRADY and B. PEARLMUTTER. Hard-lost: Modified k-means for oriented lines. Proceedings of the Irish Signals and Systems Conference, 2004.

[12] P. O'GRADY, B. PEARLMUTTER, and S. RICKARD. Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*, 15(1), 2005.