

# Fetal Heart Rate Deceleration Detection using a Discrete Cosine Transform Implementation of Singular Spectrum Analysis

Philip A. Warrick<sup>1,4</sup>, Doina Precup<sup>2</sup>, Emily F. Hamilton<sup>3,4</sup>, and Robert E. Kearney<sup>1</sup>

**Abstract**— We present a method for decomposing a signal into near-orthogonal components via the discrete cosine transform (DCT) and apply this in a novel way to change-point detection based on singular-spectrum analysis (SSA). The algorithm is applied to fetal heart rate (FHR) monitoring to improve the detection of deceleration events.

## I. INTRODUCTION

FHR monitoring provides an important diagnostic signal for the assessment of intra-partum fetal condition. The FHR manifests several patterns that have postulated physiological significance (e.g. variability, accelerations and decelerations) and visual recognition of these patterns is an important skill of an obstetrician. Signal analysis of the FHR can be used to automate detection of these patterns although the presence of noise and artifact pose significant challenges to success. This study will focus on the deceleration event, a momentary decrease in the FHR that may last from 15s to 5 min. Detection of these events is complicated by their varying length, morphology, and degrees of background variability as well as the distracting presence of artifactual events.

Signal decomposition into orthogonal components is a standard approach for generating compact representations of non-stationary signals. Singular-spectrum analysis (SSA) is one such method that has been known since the original paper of Broomhead and King [1], but which has only recently been applied in a sequential manner suitable for on-line analysis [2]. The SSA technique uses the Karhunen-Loève transform (KLT) to determine the main structure of the underlying signal. By restricting the signal representation to the eigenvectors containing the majority of the signal energy, a compact signal approximation can be extracted.

As their name implies, change-point detection algorithms attempt to detect the location of changes in the characteristics of time-series data. Many such algorithms exist, but they have often been applied to simple parametric models [3]. In [2] an on-line form of the non-parametric SSA change-point detection algorithm is presented. The SSA produces an  $L$ -dimensional hyperplane from  $M$ -dimensional time-series data ( $L \leq M$ ); the distance of the  $M$ -dimensional vectors generating the model (and which form the correlation matrix

R) projected onto the  $L$ -dimensional hyperplane should be small. The change-point detection is obtained by projecting recent data (i.e. data not used to generate the model) and applying a decision rule to the distance from the hyperplane.

Electro-cardiogram (ECG) ST-segment analysis has used KLT in numerous settings to assess cardiac condition. In [4] the KLT was used to characterize the ST segment of 2-channel ECG signal to detect ischemic episodes. They use the first five KL coefficients as a feature vector whose trajectory over time is monitored with reference to a mean feature vector. Stepwise transitions in the trajectory are considered normal while smooth changes indicate a possible ischemic ST episode. In [5] the T-wave is characterized over time by the first KL coefficient to detect sleep apnea.

This paper will briefly present the key concepts underlying KL decomposition. An efficient discrete cosine transform (DCT) implementation of the on-line SSA algorithm will be described and applied to the classification of deceleration candidate events from a database of FHR tracings (see Figure 1). To the knowledge of the authors, no such DCT-based SSA implementation has been reported in the literature.

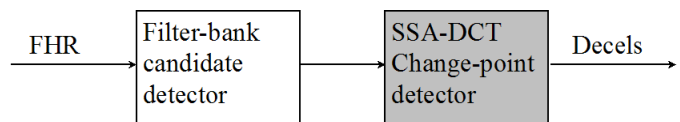


Fig. 1. Block diagram of overall processing. The second shaded block is the focus of this paper.

## II. BACKGROUND

### A. KL decomposition

In general, given any set of orthonormal basis functions  $\phi_n$  that span a linear vector space of dimension  $N$ , any function  $f(k)$  in that space can be represented as [6]:

$$f(k) = \sum_{n=0}^{N-1} \theta_n \phi_n(k), \quad 0 \leq k \leq N-1 \quad (1)$$

where the spectral coefficients  $\theta_j$  are given by the inner product

$$\theta_j = \sum_{k=0}^{N-1} f(k) \theta_j^*(k), \quad 0 \leq j \leq N-1 \quad (2)$$

It is a well-known result that the KLT finds the set of basis functions that best represent the signal in a mean-squared error sense. In addition, the decomposition of the signal with these basis functions is such that 1) the mean-square error of

Manuscript submitted May 23, 2005. Corresponding author: Philip Warrick (e-mail: philip.warrick@mcgill.ca). The authors acknowledge the financial support of this work by LMS Medical Systems, Inc.

<sup>1</sup>Biomedical Engineering Department, McGill University, Montreal, Quebec, Canada. <sup>2</sup>School of Computer Science, McGill University, Montreal, Quebec, Canada. <sup>3</sup>Department of Obstetrics and Gynecology, McGill University, Montreal Quebec, Canada. <sup>4</sup>LMS Medical Systems, Inc., Montreal, Quebec, Canada.

a truncated representation is also minimized, 2) it has optimal energy compaction: it contains the most variance (energy) in the fewest number of transform coefficients and 3) it minimizes the total representation entropy of the sequence. The result of the minimization criterion is that:

$$(\mathbf{R} - \lambda_i \mathbf{I}_N) \phi_i = 0, \quad 0 \leq i \leq N - 1 \quad (3)$$

where  $\mathbf{R} = \mathbf{E}[\mathbf{f}\mathbf{f}^T]$  is the covariance matrix of the input signal  $f$ . This is called the eigenvalue problem where  $\lambda_i$  and  $\phi_i$  are the eigenvalues and the eigenvectors of  $\mathbf{R}$ , respectively. The truncation error of reconstruction is minimized when the eigenvalues are ranked in descending order.

The drawback of the KLT is that the basis functions are signal dependent and therefore cannot be predetermined. In addition, finding the eigenvalues of  $\mathbf{R}$  is a computationally intensive task, especially as  $N$  becomes large. To overcome these impracticalities, the discrete cosine transform (DCT) is often a viable alternative as it provides a good approximation to the KLT for many signals, while using a fixed set of basis vectors. The DCT transform basis vectors are defined as:

$$\theta_k = \frac{1}{c_r} \cos \frac{(2k+1)r\pi}{2N}, \quad 0 \leq k, r \leq N - 1, \quad (4)$$

$$\text{where } c_r = \begin{cases} \sqrt{N} & r = 0 \\ \sqrt{\frac{N}{2}} & r \neq 0 \end{cases}$$

The eigenvalue approximation from the calculation of DCT coefficients over  $n$  signal realizations of length  $N$  is given by the average of the squared DCT coefficients [7]:

$$\lambda_j(n) = \frac{1}{n} \sum_{i=1}^n C_j^2(i), \quad j = 0, 1, \dots, N - 1 \quad (5)$$

where  $C_j$  is the  $j$ -th DCT coefficient. It has been shown that for a stationary zero-mean, first-order Markov process the DCT is asymptotically equivalent to the KLT as the sequence length increases and as the adjacent correlation coefficient  $\rho$  tends to unity [7]. The DCT is therefore near optimal for many correlated signals encountered in practice and fast algorithms exist for its calculation; indeed it is the industry-standard in image and speech transform coding [6].

### B. On-line SSA change-point detection

For a time series  $x_1, x_2, \dots, x_N$ , let  $M$  be an integer called the lag, where  $M \leq \frac{N}{2}$  and let  $K = N - M + 1$ . Define the  $M \times K$  Hankel matrix

$$\mathbf{X} = [X_1 X_2 \dots X_K] = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & \dots & x_N \end{pmatrix} \quad (6)$$

where  $x_{ij} = x_{i+j-1}$ .  $\mathbf{X}$  can be considered as multivariate data with the  $M$  rows as characteristics and the  $K$  column vectors  $X_j$  as observations. Defining the  $M \times M$  lag-covariance matrix  $\mathbf{R} = \mathbf{X}\mathbf{X}^T$  and calculating its singular value decomposition (SVD), we determine the orthonormal eigenvectors  $U_i$  and eigenvalues  $\lambda_i$ ,  $i = 1, 2, \dots, M$  such

that  $\mathbf{R}\mathbf{U} = \mathbf{U}\mathbf{D}$ , where  $\mathbf{D}$  is a diagonal matrix with  $\lambda_i$  along the diagonal and  $\mathbf{U} = [U_1 U_2 \dots U_M]$ .

The magnitude of the eigenvalue  $\lambda_i$  is equal to the variance in  $\mathbf{R}$  spanned by (or energy content of) the eigenvector  $U_i$  [6]. In our case, given the orthonormal set of eigenvectors  $U_i$ , reconstruction of  $X_i$  is possible by summing the product of the spectral coefficients (the inner products of  $X_i$  and  $U_k$ ) and the eigenvectors:

$$X'_i = \sum_{k=1}^L (X_i \cdot X_k) U_k, \quad 0 \leq L \leq M \quad (7)$$

If  $L = M$  we have perfect reconstruction while for  $L < M$  approximations to  $X_i$  are formed using the  $L$  eigenvectors having the highest eigenvalues. The assumption made by the SSA algorithm is that given an appropriate choice of  $L$ , the reconstruction will tend to correspond to signal while the residue energy (for the unused terms  $L + 1, \dots, M$ ) will correspond to noise. The residual energy  $e_r$  is calculated as the difference in Euclidean norms:

$$e_r(X_i) = \|X_i\|^2 - \|X'_i\|^2 \quad (8)$$

The SSA algorithm for change-point detection calculates the basis functions over a base interval and evaluates the residual energy of the reconstruction of another interval called the test interval. The intervals are shifted in time across available samples and the residual energy is calculated for each shift. Specifically, the base and test intervals are used to construct corresponding base and test matrices as follows. For time shift  $n$ , the  $\mathbf{X}$  matrices have the following form:

$$\begin{pmatrix} x_{n+1} & x_{n+2} & \dots & x_{n+K} \\ x_{n+2} & x_{n+3} & \dots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \dots & x_{n+N} \end{pmatrix} \quad (9)$$

K-vector base matrix

$$\begin{pmatrix} x_{n+P+1} & x_{n+P+2} & \dots & x_{n+P+Q} \\ x_{n+P+2} & x_{n+P+3} & \dots & x_{n+P+Q+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+P+M} & x_{n+P+M+1} & \dots & x_{n+P+Q+N} \end{pmatrix} \quad (10)$$

Q-vector test matrix

where  $P$  is the position of the test interval relative to the beginning of the base interval and  $Q$  is the number of observations taken from the test interval.

The difference metric at sample  $n$  is given by the summation of the residual energies of the test matrix vectors

$$D_n = \sum_{i=n+P+1}^{n+P+Q} e_r(X_i) \quad (11)$$

where the basis functions  $U_i$  for the reconstruction are calculated from the base matrix. This difference metric is then adjusted by an estimate of the underlying noise energy, taken as the difference metric of the base matrix where

the last null hypothesis  $\mathbf{H}_0$  (no change) occurred: i.e.  $m$  is chosen as the largest value of  $m \leq n$  under  $\mathbf{H}_0$ :

$$S_n = \frac{D_n}{v_n} \quad \text{where} \quad v_n = \sum_{i=m+1}^{m+K} e_r(X_i) \quad (12)$$

If  $v_n$  is considered an estimate of the variance of the noise at sample  $m$  (where it is assumed that there is very little signal energy),  $S_n$  normalizes the distance metric such that it is not depend on this (assumed unknown and possibly non-stationary) quantity.

It is difficult to select a threshold for the change statistic that could apply to an arbitrary sequence. Nonetheless, under some simplifying assumptions, it is shown in [2] that the following broad rule applies: when the distribution of the change statistic  $D_n$  is assumed to be Gaussian, a threshold can be chosen such that the probability of an event under  $\mathbf{H}_0$  is approximately  $\alpha$ , where  $\alpha > 0$  is a significance level. In [2] the correspondence of  $\alpha$  to a threshold  $h$  of the  $S$  statistic is elaborated, and we use this  $h$  calculation in the subsequent discussion.

### C. FHR database and candidate detector

The database for this study consisted of 15 FHR tracings containing 535 decelerations as marked by an expert obstetrician. We have described in [8] an approach that uses filtering to detect deceleration candidates. The filtering stage consists of a bank of band-pass filters responsive to decelerations over various overlapping frequency ranges. Candidate events, whose extent corresponds to the zero-crossings of a particular band-pass filter, are placed in a competition with overlapping events generated by the other band-pass filters, and only the highest amplitude events survive. This detector detects 528 of the 535 decelerations of the study database, although it also detects 1285 false positives. One of the goals of this study is to improve on the discrimination of the candidate detector.

## III. METHODS

### A. SSA Change Point Detection: Block DCT implementation

While the DCT implementation provides some reduction of the processing load compared to the KLT, further improvements were made by using a block implementation of the DCT version of the algorithm. Because the DCT eigenvalue approximations are summations of the DCT coefficients of the observation vectors (eqn. 5), a recursive calculation at each time sample is possible [7]:

$$\lambda_j(n) = \lambda_j(n-1) + \frac{1}{N} (C_j^2(n) - C_j^2(n-N)), \quad (13)$$

$$j = 0, 1, \dots, N-1$$

Performing the DCT calculations in block fashion produced a more efficient SSA implementation. For one typical FHR segment of approximately 2400s duration, a 40-fold decrease in the SSA processing time was observed with the block implementation (from 1000s to 25s) compared to repeated and highly redundant DCT calculations at one-sample time shifts of the analysis block.

Using several calibration FHR tracings, we observed good agreement between the KL and DCT implementations of the SSA algorithm (see one example in Figure 2). Some higher frequency spikes are better resolved by the KLT, but the responses are otherwise similar.

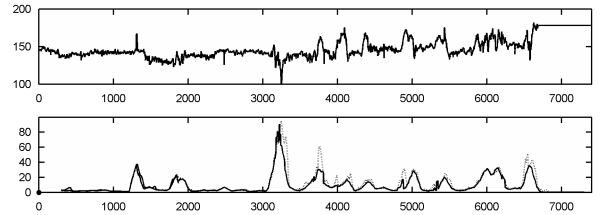


Fig. 2. The lower plot compares the KLT (dashed) and DCT (solid) implementations of the SSA statistic  $S$ . The input FHR appears above.

### B. SSA Parameter selection

The SSA algorithm is quite parameter-sensitive. Perhaps the single most important parameter for selection is the observation length  $M$ . Our preliminary investigations suggested that  $M$  should be chosen to be of similar duration to the event of interest. The significance level of the threshold  $h$  for the SSA  $S$  statistic was set to values close to  $\alpha = 0.1$  (as suggested in [2]). The number of observations  $K$  needs to be large enough to create a relatively stable base interval representation but not so long as to smooth out distinct regions of quasi-stationarity [2]. The other main choice is the  $P$  value: whether to place the test interval within the base interval or following it. Embedding the test interval in the base interval has the advantage of including the before- and after-event signal in the determination of the base matrix, which may be appropriate for the typical deceleration event that is preceded and followed by some common steady state. On the other hand, placing the test interval after the event should accentuate the difference between the base and test intervals since no test signal is included in the base matrix calculation. We used the latter approach in the experiments that follow. The length of the test interval  $Q$  was chosen to be somewhat smaller than  $M$  (e.g.  $Q = \frac{M}{4}$ ) and the number of eigenvectors  $L$  was kept as small as possible (e.g.  $L = 5$ ) since often the most important changes occur in the highest energy components.

### C. SSA on candidate detected FHR

Using the location of previously detected deceleration candidates focused the analysis by matching the  $M$  value to the event duration and reducing the search space significantly. Given the set of candidates, local SSA was done in the vicinity of each candidate, using the candidate length for the SSA observation length  $M$  in each case (i.e. as an estimate of the true event length). For larger values of  $M$ , the processing burden was excessive and for this reason the data was decimated. The decimation factor was determined by the ratio of the event length  $L_C$  to the minimum event length  $L_M$  of 15s, i.e. the decimation factor is defined as  $D = \text{round}(2L_C/L_M)$ , where  $\text{round}(\cdot)$  denotes rounding

	TP In	FP In	TP Missed	FP Reduction
Standard SSA	528	1285	64	279
Base-hold SSA	528	1285	49	278

TABLE I  
SSA-DCT DECELERATION DETECTION RESULTS

to the nearest integer. With the minimum candidate length restricted to 15s, the number of decimated samples  $L_D = \text{round}(L_C/D)$  falls into the range of 25 to 37 samples. Thus the results of the analysis are relatively scale-invariant over a broad range of event durations. The SSA analysis begins and ends at fixed factors of  $M$  before and after the event, chosen in a way to give sufficient context to the analysis of the event. By experimentation, we found that factors of 16 and 8, respectively, provided enough context without incurring excessive decision delay.

#### D. Base-hold SSA

The standard SSA algorithm encounters difficulty in intervals where repeated decelerations occur in close proximity. Under these conditions the first events of the series may be recognized, but subsequent events are not since the base and test intervals tend to contain similar energy (i.e. of morphologically similar decelerations) as analysis progresses forward in time. To address this, we modified the algorithm to hold the base interval constant when changes are detected until the next null hypothesis  $\mathbf{H}_0$  is encountered.

For these conditions, Figure 3 illustrates how holding the SSA base interval to the last  $\mathbf{H}_0$  interval can improve detection. The standard SSA base interval  $B_S$  contains an event, reducing its distinction from the test interval  $T$ . The base-hold interval  $B_H$ , having less low-frequency energy, is retained as the base interval for the successive event candidates, and the energy difference between  $B_H$  and  $T$  is sufficient to signal a change-point.

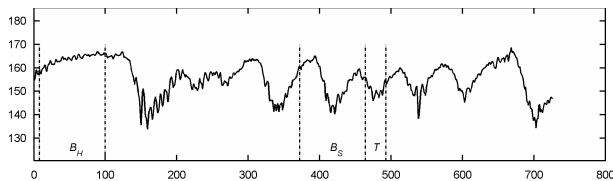


Fig. 3. Example base-hold. For the sixth event of Figure 4, the locations of the base intervals for the standard and base-hold SSA are shown between dotted vertical lines as  $B_S$  and  $B_H$  respectively. The test interval is indicated by  $T$ .

## IV. RESULTS

The results of the detection performance of the standard and base-hold SSA are shown in Table I. The true and false positives (TP, FP) of the candidate detector are the input to the two variants of the change-point detector. Both detectors reduced the number of false positives by approximately 21.6% (279 or 278 of the original 1285). However, the base-hold approach had 23% fewer misses (49 vs. 64). Thus for

equivalent false-positive rates, the base-hold approach shows improved sensitivity over that of standard SSA.

Figure 4 shows an example of repetitive deceleration patterns, which are problematic for standard SSA. The base-hold SSA algorithm detects two decelerations that are missed by standard SSA. Both approaches have the same false positives.

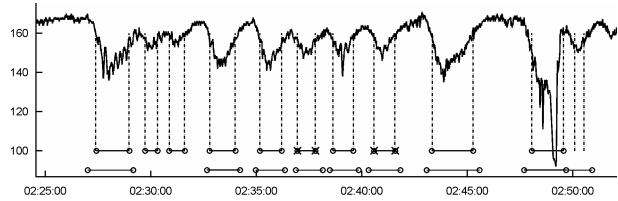


Fig. 4. Example of a region of repetitive decelerations. The dotted vertical lines indicate the extents of candidate detector markings. The lower and upper horizontal bars indicate the extents of the expert markings and SSA markings, respectively. The standard SSA algorithm misses the sixth and eighth decelerations (marked by  $\times$ ) whereas the base-hold approach detects them. Both approaches have the same false positives (the second and third events).

## V. DISCUSSION

The standard SSA algorithm has been used successfully in many contexts where changes are infrequent [2]. Our efficient DCT implementation of SSA is of general applicability to these problems, especially where the processing data volume is large and the input signal is sufficiently correlated. In FHR analysis, the assumption of infrequent changes does not always apply and our modifications improve the resolution of single events within a series of similar events. While this has shown to improve overall results on our dataset, it is also conceivable that a series of similar non-events can be wrongly classified as a group as a result of this modification.

Rather than using this change-point detector in isolation in our FHR analysis, we are investigating the incorporation of the SSA statistic into an ensemble of other features within a more comprehensive classifier.

## REFERENCES

- [1] D. Broomhead and G. King, "Extracting qualitative dynamics from experimental data," *Physica D*, vol. 20, pp. 217–236, 1986.
- [2] V. Moskvina, "Application of the singular spectrum analysis for change-point detection in time series," Ph.D. dissertation, Cardiff School of Mathematics, Cardiff University, 2001.
- [3] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall, 1993.
- [4] F. Jager, R. Mark, G. Moody, and S. Divjak., "Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève transform," in *Computers in Cardiology*, 1992, pp. 691–694.
- [5] C. Maier, H. Dickhaus, M. Bauch, and T. Penzel, "Comparison of heart rhythm and morphological ECG features in recognition of sleep apnea from the ECG," in *Computers in Cardiology*, A. Murray, Ed., 2003, pp. 311–314.
- [6] A. Akansu and R. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, Wavelet*, 2nd ed. Englewood Cliffs, New Jersey: Academic Press, 2001.
- [7] S. Haykin, *Adaptive Filtering*, 4th ed. Englewood Cliffs, New Jersey: Prentice-Hall, 2002.
- [8] P. Warrick, E. Hamilton, and M. Macieszczak, "Neural network based detection of fetal heart rate patterns," in *IJCNN05*, 2005, in press.