

Fetal Heart Rate Deceleration Detection from the Discrete Cosine Transform Spectrum

Philip A. Warrick^{1,4}, Doina Precup², Emily F. Hamilton^{3,4} and Robert E. Kearney¹

Abstract—Automated detection of decelerations in fetal heart rate (FHR) signals can be posed as a problem of signal detection in the presence of noise. We present an algorithm that adaptively selects the resolution of analysis and uses the discrete cosine transform (DCT) to describe the spectrum at short-term and longer-term scales. In so doing we generate near-orthogonal and scale-invariant features that are presented to a feedforward neural network for classification.

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. baseline variability, accelerations and decelerations) and visual recognition of these patterns is an important skill of an obstetrician.

Feature detection is the process of segmenting the FHR into three constituent parts: baseline, acceleration and deceleration. FHR baseline is a region of the signal where the heart rate is relatively constant. Regions where there are significant excursions above and below the baseline are termed accelerations and decelerations, respectively. Their duration is typically in the range of 15s to several minutes and they have a minimum amplitude of 15bpm (beats per minute). These are to be distinguished from the normal baseline variability whose excursions are continuous and relatively smaller. This study focuses on the deceleration event, a momentary decrease in the FHR. Detection of these events is complicated by their varying length, morphology, and degrees of background variability as well as the distracting presence of artifact (see Figure 1). FHR is normally measured in tandem with the maternal uterine pressure (UP) and the primary physiological mechanism responsible for decelerations is umbilical-cord compression during uterine contractions.

Signal decomposition into orthogonal components is a standard approach for generating compact signal representations and the Karhunen-Loève transform (KLT) gives the theoretically optimal such representation. By

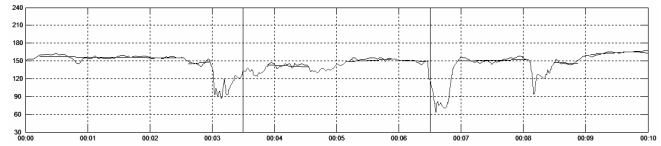


Figure 1: Example plot of four FHR decelerations of varying morphology and the baseline segments between them. The second deceleration is more difficult to distinguish from the ambient variability.

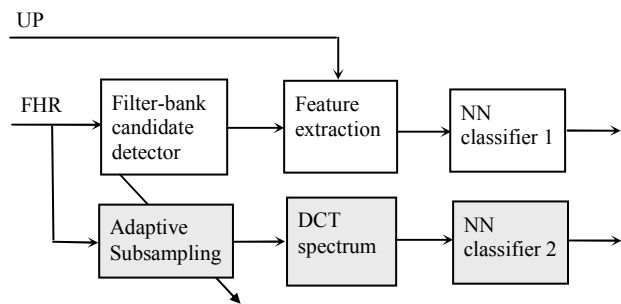


Figure 2: Block diagram of the deceleration detection schemes for the reference (manually selected features, top pathway) and proposed (features from the DCT spectrum, bottom pathway, shaded in gray) approaches.

restricting the signal representation to the eigenvectors containing the majority of the signal energy, a compact signal approximation can be extracted. We use a discrete-cosine transform (DCT) to approximate the KLT on selected segments of the FHR and use these features as inputs to classify candidate decelerations with a neural network.

This paper presents the key concepts underlying KLT and DCT decomposition and selectively surveys some current signal-detection methods using these techniques. An application to FHR analysis is described in which measurement features obtained from the DCT decomposition of subsampled FHR are presented to a feedforward neural network for classification of deceleration events. When compared to known event locations marked by an expert obstetrician, the detection results compare favourably to a neural network classifier that uses manually selected features. The two approaches are illustrated in Figure 2.

II. BACKGROUND

A. DCT approximation to KL decomposition

In general, given any set of orthonormal basis functions ϕ_n that span a linear vector space of dimension N , any function $f(k)$ in that space can be represented as [1]:

Manuscript submitted Apr. 22, 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.

¹Department of Biomedical Engineering, McGill University, Montreal, Quebec, Canada.

²School of Computer Science, McGill University, Montreal, Quebec, Canada.

³Department of Obstetrics and Gynecology, McGill University, Montreal Quebec, Canada.

⁴LMS Medical Systems, Inc., Montreal, Quebec, Canada.

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

where the spectral coefficients θ_j are given by the inner product

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

It is well known that the KLT finds the set of basis functions that best represent the signal in a mean-square error sense. In addition, the decomposition of the signal with these basis functions is such that the mean-square error of a truncated representation is also minimized. The result of this minimization criterion is that

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

where $\mathbf{R} = E[\mathbf{f}\mathbf{f}^T]$ is the covariance matrix of the input signal f . This is called the eigenvalue problem where λ_i and ϕ_i are the eigenvalues and the eigenvectors of \mathbf{R} , respectively. The truncation error of the reconstruction is minimized when the eigenvalues are ranked in descending order. Other noteworthy optimal properties of this transform are that 1) it has optimal energy compaction: it contains the most variance (energy) in the fewest number of transform coefficients and 2) it minimizes the total representation entropy of the sequence.

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 DCT is often a viable alternative that approximates well the KLT for many signals, while using a fixed set of basis vectors. The DCT basis vectors are defined as:

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

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

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 [3]:

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

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 ρ tends to unity [3]. 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 [1].

B. Previous Work

We have described in [7] an approach that uses filtering, feature extraction and neural-network classification to detect decelerations. 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 virtually all true decelerations at the expense of approximately three false alarms for every detection: for a database of events this detector had a sensitivity of 99.0%, and a positive-predictive value (PPV) of 25.4%. PPV indicates how often an example classified as positive is in fact positive; it is defined as $PPV = TP/(TP+FP)$, where TP and FP refer to true and false positives, respectively. The database consisted of 161 FHR tracings (762 hrs) containing 5490 decelerations as marked by an expert obstetrician.

Based on consultation with the expert, 20 features deemed important to the classification task were extracted for all candidates. These included the event duration, mean and standard deviation of FHR and high-frequency energy (variability) before during and after the event, maximum slopes, area, mean and maximum height, and proximity of contraction. These features were used to train a feed-forward neural network to which we shall refer henceforth in this paper as the reference classifier. This detection approach has been compared favourably to other deceleration detectors in the literature in [7].

Electro-cardiogram (ECG) ST-segment analysis has used KLT in numerous settings to assess cardiac condition. In [4] the KLT is 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. They report better detection with this approach than by using the R-R interval (equivalent to the FHR) directly.

III. METHODS

A. Candidate Event selection

The problem of distinguishing decelerations from background variability can be posed as that of signal detection in the presence of noise. Given the continuous nature of the variability, it is advantageous to restrict the search space to that of the most likely candidate decelerations. We did this using the location of deceleration candidates generated by a detector described in [7].

B. Local and Neighbourhood Spectra

With this set of candidates, the approach was to estimate

the local spectrum of the event and the spectrum of the recent past to near future using DCT decomposition. However, given the wide range of possible event durations, it is important to describe the spectrum in a scale-invariant manner for subsequent analysis classifier. To achieve this, we reduced the sampling frequency of the FHR signal (originally sampled at 4Hz) in accordance with the event duration and analyzed the spectrum of this decimated signal. The decimation factor was determined by the scale factor of the candidate length L_C compared to the minimum event length L_M of 15s: the (integral) decimation factor is defined as $D = \text{round}(2L_C/L_M)$, where $\text{round}(\cdot)$ denotes rounding 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. Alternatively, we could have fixed the number of samples for each candidate and interpolated, but restricting the decimation to integral factors permits the use of finite-impulse-response (FIR) digital filters, which is advantageous because of their efficient FFT implementation and their desirable linear phase properties. We widened the region including the samples used to calculate the local spectrum by 25% on each end to increase the number of signal realizations to $n_L = L_D/2$ (of length $N = L_D$) in the DCT eigenvalue approximation of (5).

The local spectrum alone may lack important contextual information and for this reason we also computed the spectrum of the region surrounding the event. We chose a context containing proportionally more past history ($16L_D$) than future ($8L_D$). Events near the beginning or end of the FHR signal where this context was unavailable were removed from the analysis. With N common to both segments, the number of signal realizations in this case is $n_C = (16+1+8) L_D$. Averaging over n_L or n_C permits the signal energies of corresponding local and neighbourhood coefficients to be directly compared.

The energy-compaction properties of the KLT were confirmed for the DCT approximation over a range of FHR time scales and a typical example is shown in Figure 3. These results suggest that a reduced set of DCT coefficients should provide an adequate signal estimate, assuming that the coefficients in the flat region of the graph correspond to random noise. Typically, the FHR signal energy was concentrated in a small number of coefficients (10 in the figure); hence we presumed that the spectral features should remain reasonably scale-invariant despite the variation in L_D described above. Our standard representation among all examples was to use the first $j-1$ coefficients plus a j -th component containing the residue (i.e. the sum of the coefficients from j to N).

C. Neural-network classifier

The spectral features and length L_C of each candidate in the database were applied as inputs to the training of feed-forward neural networks. Training was performed using Levenberg-Marquardt backpropagation [2], a second-order batch-training algorithm. By experimentation the “best”

neural-network architecture that we found that avoided both overfitting and underfitting was 6 input layers \times 4 hidden layers \times 1 output layer. The training was continued for 40 epochs. The experiments were performed using eight-fold cross-validation; for each fold, 16 neural networks were trained and the neural network having the lowest training error was retained. True decelerations, being fewer in number compared to the non-events, were presented three times to the network to correct the imbalance in the non-event: event proportion.

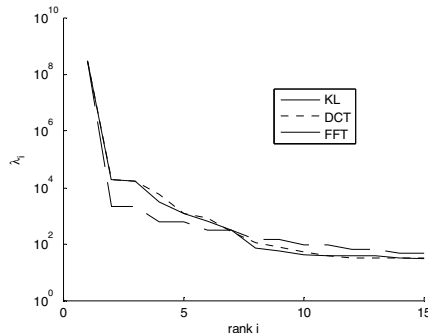


Figure 3: Comparison of highest-ranked coefficients λ_i from KLT, DCT and FFT decompositions for a typical FHR section. For KLT and DCT, the majority of the signal energy is primarily captured by the ten highest components. The close approximation of the DCT to the KLT and the reduced energy compaction of the FFT is evident.

IV. RESULTS

Receiver-operator characteristics (ROC) were generated by computing the average neural-network performance on the test data of each fold at various output thresholds. The number of spectral coefficients was varied and the best results are shown using 10 coefficients each for the local and neighbourhood spectra. The cross-validation was repeated five times for the DCT and reference classifier, using a different set of randomly selected fold combinations each time.

Figure 4 shows the average of five cross-validation results for each classifier. Here the sensitivity and false positive rates were interpolated to obtain regularly spaced sampling in order to obtain statistics at common points along the respective ROC curves. The error bars indicate the 95% confidence levels at each point. The mean area-under-curve (AUC) and their 95% confidence bounds are also shown. The DCT result is slightly better (by $\sim 1\%$) than the reference classifier in these two graphs. The closeness of the confidence intervals, however, makes it difficult to make a definitive judgment on their relative performance.

Because the ROC curve considers only the rate of false positives, it can mask the inherent tradeoff between sensitivity and the absolute number of false positives.

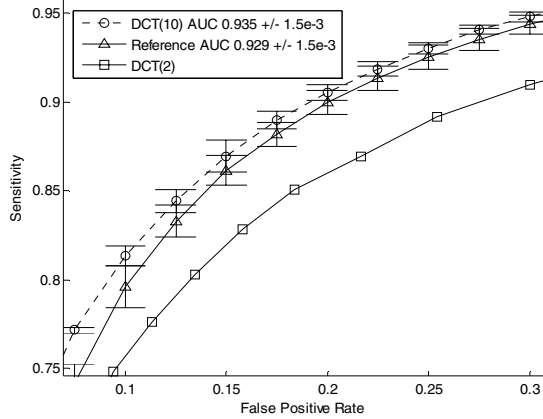


Figure 4: Average of five cross-validation results for the DCT 10 coefficients and the reference classifiers with 95% confidence intervals indicated by the error bars. The original ROC curves were resampled at regular spacing of the false positive rate in order to obtain mean and standard deviations at common points along the ROC curve of each cross validation. For reference, a very simple classifier using only 2 spectral coefficients is also shown.

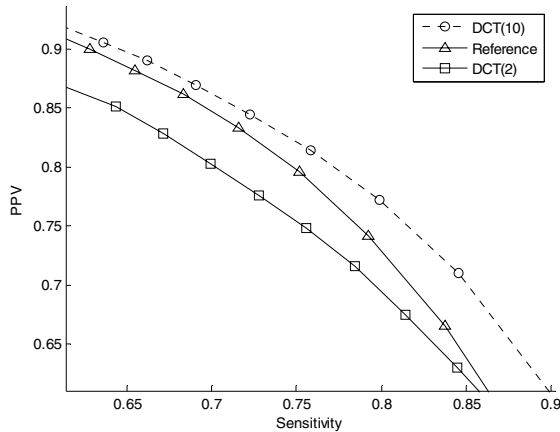


Figure 5: PPV-sensitivity (precision-recall) graphs for the three classifiers showing a slightly improved tradeoff between sensitivity and absolute number of false positives for the DCT 10 coefficient classifier.

Precision-recall (PPV vs. sensitivity) and F-measure plots better display this tradeoff and we plot these values averaged over the five simulations in Figure 5 and Figure 6. The slight improvement of the DCT over the reference classifier precision-recall graphs indicates that the absolute number of false positives at a given sensitivity is lower for the DCT approach. The F -measure is defined as the harmonic mean of the sensitivity and the PPV:

$$F = \frac{2(PPV)(Sens)}{PPV + Sens} \quad (6)$$

Its maximum value is considered to be a useful single metric of algorithm performance that combines the two measures with equal weight [6]. The maximum F -measures were slightly better for the DCT compared to the reference and the two-coefficient DCT approaches (78.5%, 77.3% and 75.2% respectively).

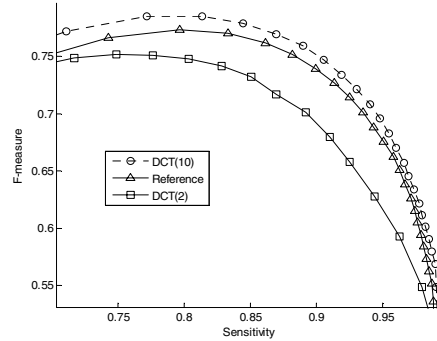


Figure 6: F-Measure vs. sensitivity for the three classifiers. The peak F-measure values for the DCT 10 coefficient, reference and DCT 2 coefficient classifiers were 78.5%, 77.3% and 75.2% respectively.

V. DISCUSSION

The DCT coefficients form a preferable feature set to the manually selected and correlated features of the reference classifier both because of their near-orthogonality and resulting reduced dimensionality, but also because they require no a-priori knowledge of what might be the “best” discriminating features. In this sense they form an unbiased set of features. Our study has shown that these benefits are achieved with performance estimates that are equal to or slightly better than those derived from the best manually selected features.

The reference classifier also had the benefit of incorporating uterine contraction information that we did not use in this study. We anticipate that incorporating this information will improve the classifier performance.

Since deceleration detection is ultimately meant to occur at the bedside in real-time, future work should consider choosing parameters that reduce decision delays. In particular, the amount of future signal used in the neighbourhood spectrum calculation should be reduced so that longer decelerations do not incur unreasonable delays. This study used conservatively large neighbourhood segments and we anticipate that such reductions will be possible.

REFERENCES

- [1] Akansu A., Haddad R., Multiresolution Signal Decomposition: Transforms, Subbands, Wavelets, Academic Press; 2nd edition, 2001.
- [2] Hagan M. and Menhaj M., Training Feedforward Networks with the Marquardt Algorithm, IEEE Transactions on Neural Networks, Vol. 5, No. 6, November 1994, pp. 989-993.
- [3] Haykin, S., Adaptive Filtering, 4th ed., Prentice-Hall, New Jersey, 2002.
- [4] Jager F., Mark R.G., Moody G.B., Divjak S., Analysis of Transient ST Segment Changes During Ambulatory Monitoring Using the Karhunen-Loève Transform, Computers in Cardiology 1992, pp. 691-694.
- [5] Maier C., Dickhaus H., Bauch M., Penzel T., Comparison of Heart Rhythm and Morphological ECG Features in Recognition of Sleep Apnea from the ECG. In: A. Murray (eds.), Computers in Cardiology 2003. IEEE Press; pp. 311-314; 2003.
- [6] van Rijsbergen, C.J., Information Retrieval. Butterworths, London, 1979.
- [7] Warrick P.A., Hamilton E., Macieszczak M., Neural Network Based Detection of Fetal Heart Rate Patterns, IJCNN05 Conference, in press, 2005.