Discriminative Ensemble Averaging for Accelerated Denoising

Corey McCall

Abstract

It is difficult to filter a signal with high-amplitude in-band noise if an independent noise reference cannot be obtained. If the signal is periodic relative to a known sequence of temporal markers, Ensemble Averaging can be used to reduce the noise by averaging the signal over several periods in the same way that artificial neural networks average many modest experts to achieve a higher overall accuracy. Although this method effectively reduces in-band noise for this type of signal, it costs several time periods of waiting and is subject to corruption by high-amplitude spikes that take take several periods to reduce. In this paper, we use *k*-means clustering to discriminate against corrupted periods in order to achieve faster convergence of the ensemble average in a ballistocardiography dataset recorded on NASA's Zero-G aircraft. By using a single clean training example from each subject to select the least noisy cluster, ensemble averaging is accelerated for all ten subjects in the dataset within the first 4 cycles.

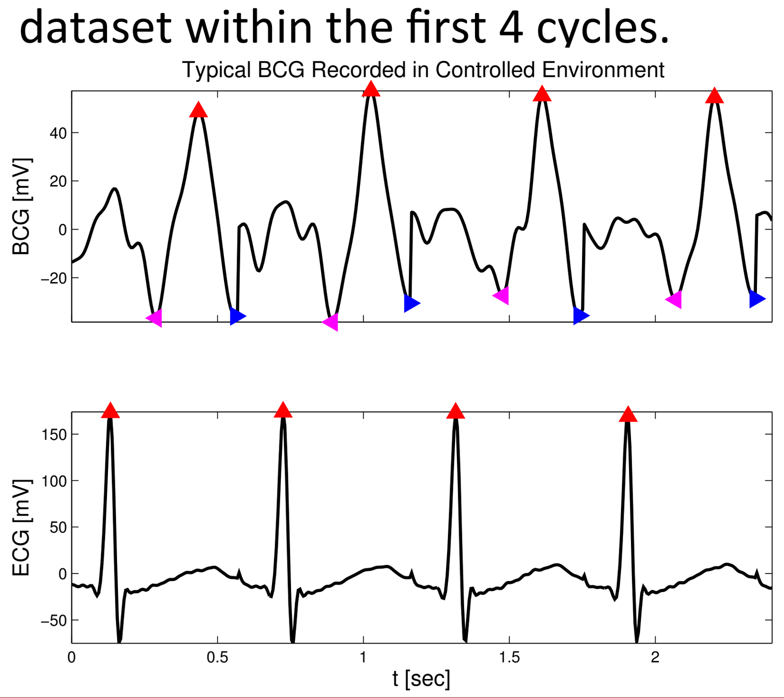


Figure 1. In a controlled environment, BCG can be easily obtained by standing still on a weight scale as shown (top). Various useful physiological parameters can be derived from the timing differences between the simultaneously recorded ECG (bottom) R peaks (red triangles), and the IJK complex of the trailing BCG waveform marked by the magenta (I), red (J), and blue (K) triangles.

Introduction

A discriminative Ensemble Averaging (EA) approach is described for measuring small signals with high- amplitude in-band noise. The motivating example is measuring the c(BCG) of astronauts onboard a crowded space capsule. The signal can be ensembled by synchronizing with the simultaneously obtained electrocardiogram (ECG), in which the ECG R peaks proceed each BCG waveform complex. Although each BCG beat can be effectively referenced to an ECG timing marker, the congested setting induces frequent high-amplitude motion noise spikes from collisions with weightless objects and other astronauts, each of which increases the convergence time of the average. In order to solve this issue, the convergence is accelerated by discriminating against corrupted beats. This is done by clustering each sample of the ensemble at each step in order to filter out abnormal samples that negatively affect the average and increase the convergence time. Unlike naive EA, which blindly averages each beat together at each sample, the algorithm described here constructs the average from a subset of of the recorded beats determined at each sample by clustering.

BCG Dataset

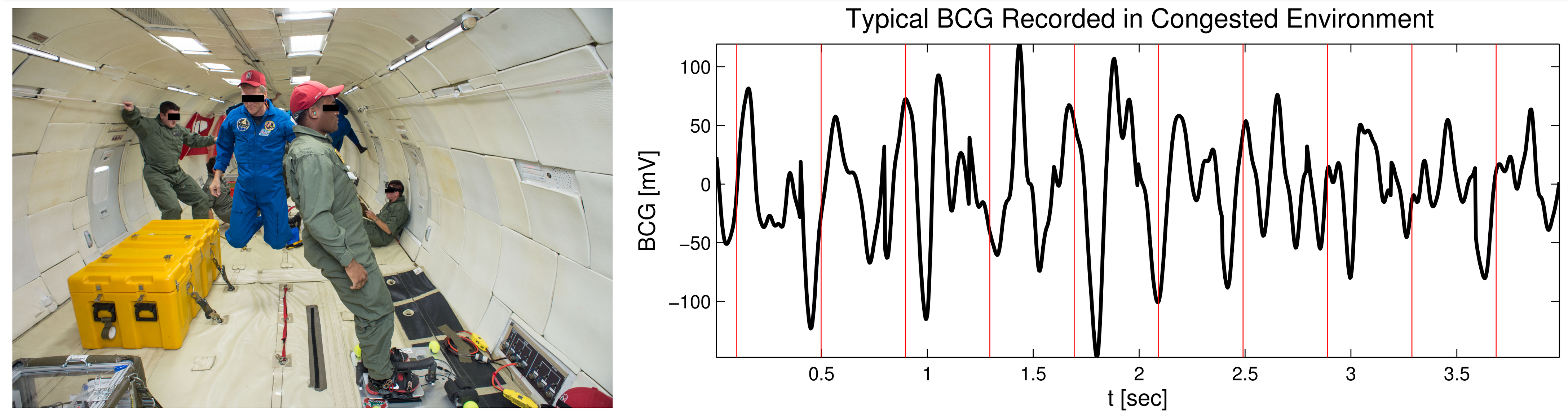
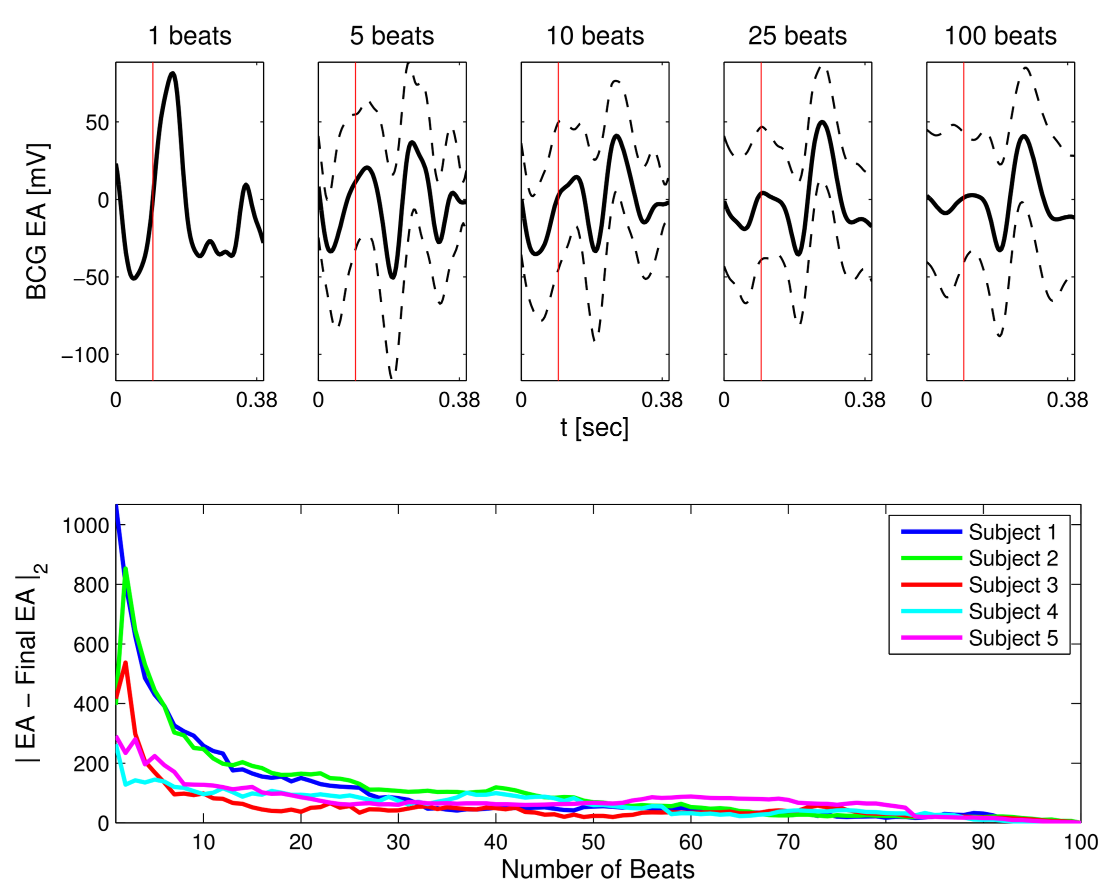


Figure 2. In a congested environment, BCG is difficult to distinguish because of the in-band noise that is much larger than the BCG component. The vertical red bars indicate ECG R peaks used to synchronize the ensemble.

Ensemble Averaging



$$\mathbf{X} = \begin{bmatrix} \text{———} & (x^{(1)})^T & \text{———} \\ \text{———} & (x^{(2)})^T & \text{———} \\ & \vdots & \\ \text{———} & (x^{(m)})^T & \text{———} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$\hat{x}^* \triangleq \frac{1}{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = a\}} \mathbf{X}^\top \mathbb{1}\{\mathbf{Y} = a\}$$

Figure 3. EA convergence for 5 subjects calculated naively by averaging each beat in sequence. The top row shows an example for Subject 1 as the average converges. The vertical red bars indicate the ECG R peaks used to synchronize the ensemble, and the thin dashed lines indicate the standard deviation of the average at each point.

Discriminative Ensemble Averaging

Algorithm 1 Discriminative Ensemble Averaging

Input: $X \in \{x^{(i)} \in \mathbb{R}^n; i = 1, \ldots, m\}$, $x_{train} \in \mathbb{R}^n$
Output: $\mu_{DEA} \in \mathbb{R}^n$
$\mu_{EA} = \mathrm{EA}(X)$
$X := X \cup x_{train}$
for $j = 1$ **to** n **do**
  $z = X_j$
  $c = \mathrm{CLUSTER}(z)$
  $z' = \{z_i; c_i = c_{m+1}\}$
  $z' := z' \setminus x_{train,j}$
  $\mu_{DEA,j} = \min(\mathrm{EA}(z'), \mu_{EA,j})$
end for

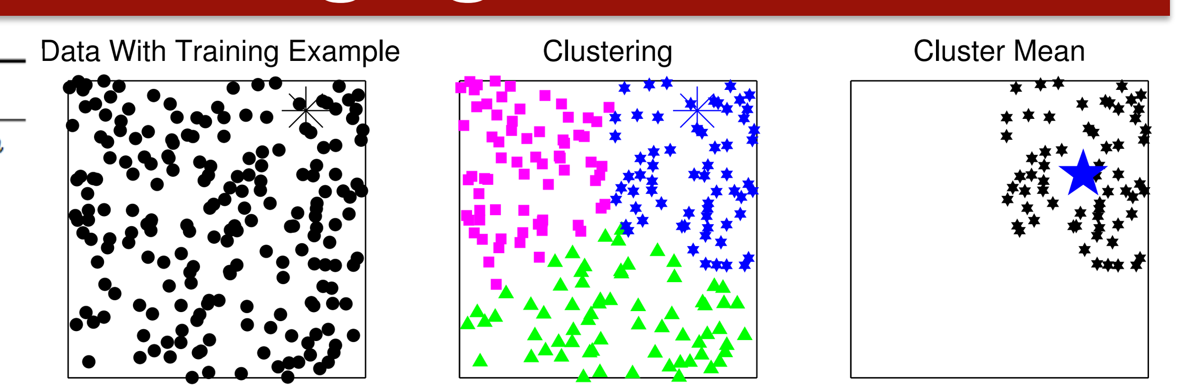
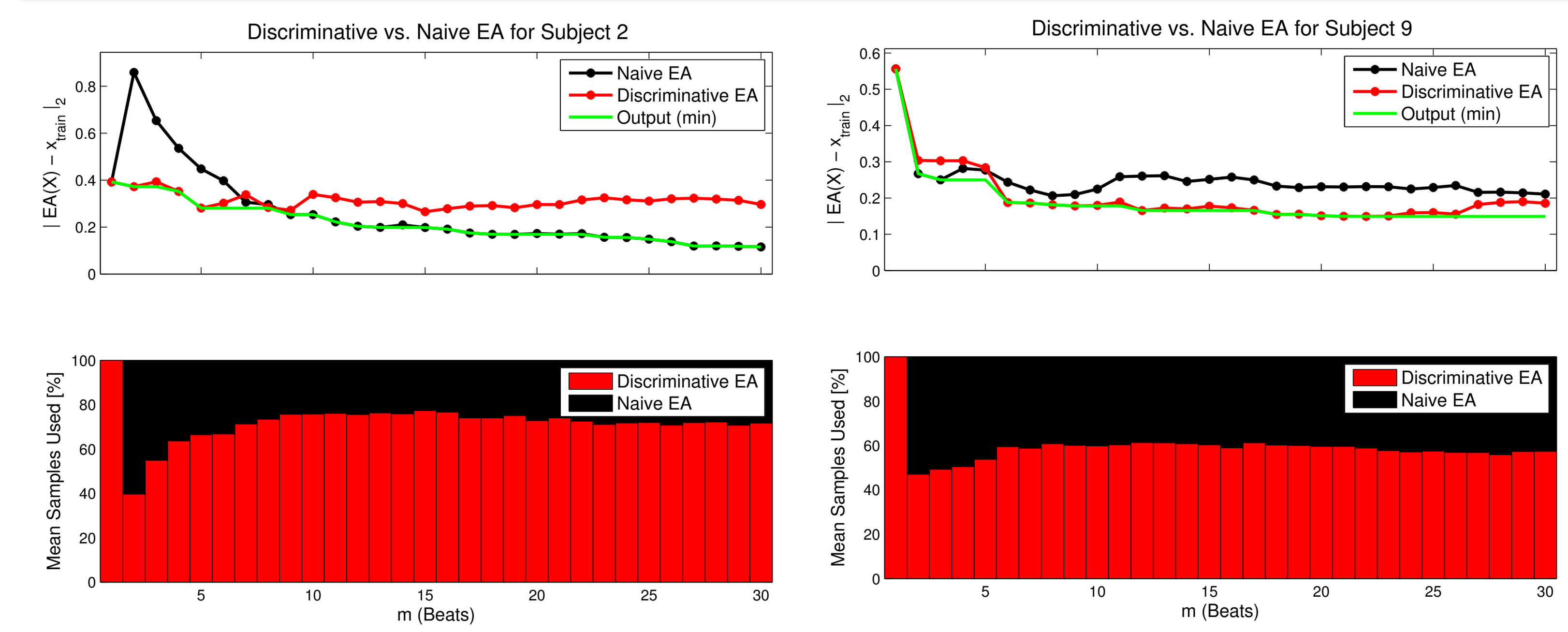


Figure 4. Graphical 2D example of Discriminative Ensemble Averaging for *k* = 3 clusters. In the left figure, a single training example is added to the dataset. In the center figure, all of the points are clustered into *k* clusters. In the right figure, the training example is removed and the mean of its cluster is taken as the discriminative EA.

Results



Figures 5 and 6. Discriminative EA increases the convergence rate compared to naive EA for typical Subject 2 (top, left) and 3 (top, right). Mean percentage of data used at each point in the corresponding discriminative EA (Bottom).

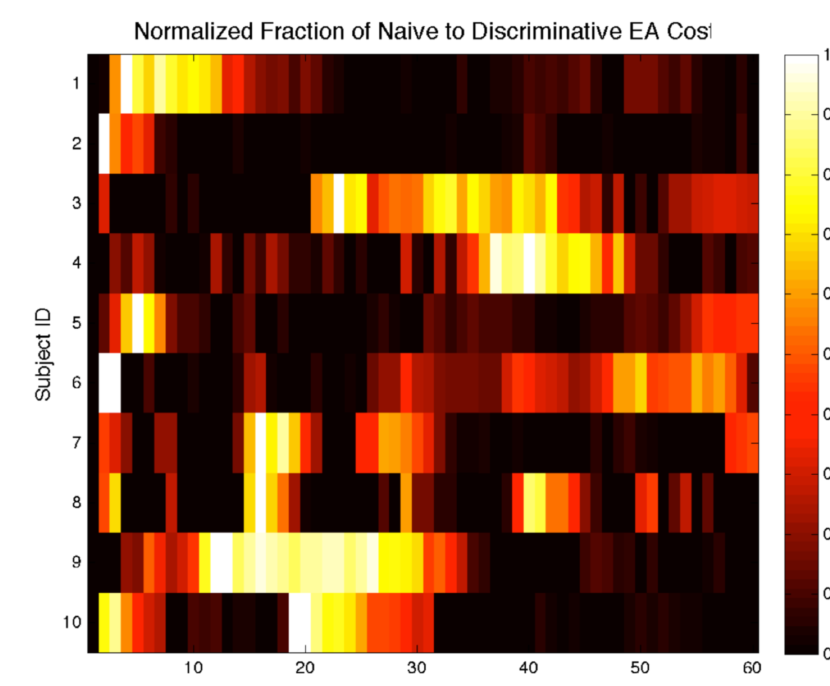


Figure 7. Algorithm performance across the entire dataset for m = 100 randomly sequenced beats. Each row is normalized to the maximum ratio of naive to discriminative EA cost functions (lighter = better). It can be seen that most subjects exhibit a performance increase in the first few cycles.

Discussion

A novel discriminative EA algorithm is described and tested on an appropriate dataset of ballistocardiogram data recorded in a very noisy environment. The algorithm, currently based on *k*-means clustering effectively increases the EA convergence rate in all subjects within 4 cycles.

Future work should include comparing this method to other non-frequency-based filtering techniques such as PCA dimensionality reduction before seeking peer-reviewed publication. It would also be valuable to test the algorithm on a dataset of simulated irregular noise for quantitative SNR analysis.

Acknowledgements

We acknowledge Professor Gregory Kovacs and Dr. Laurent Giovangrandi at Stanford University's Transducers Lab for discussing BCG signal processing techniques and sharing this dataset. We also acknowledge NASA's Reduced Gravity Office for facilitating the collection of this dataset.