# Heart Data Clustering

By Juan Gabriel Estrada Alvarez

## A refresher

- Provide a means for classification of heart conditions
- Idea: use clustering
- Problem: it is very hard to cluster heart pulse interval time series

## Why hard?

- Comparison of time series directly is not useful if they did not all begin at the same time
- Heart data taken over years
- Solution: find a time independent representation

## Solution

- Fourier Transform – map the frequency domain
- We can now use our distance measure: NRMSD
- Clustering algorithm: bottom-up

## The Data

- Learned : better if preprocessing done
  - Allow the user to provide the time independent representation of the data
  - This way the application can be generalized to any kind of data
- Here a single precision FFT algorithm was used
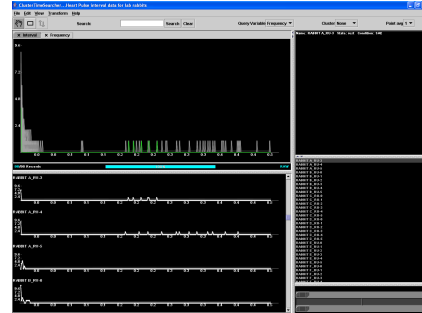
## ClusterTimeSearcher

- Based on Hochheiser's TimeSearcher app
- Takes as input a data file containing both:
  - Time series data
  - Representation of the data which will be used for clustering
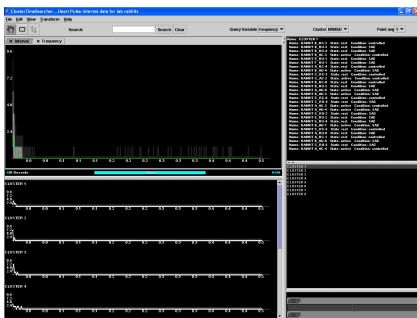- Can display the unclustered data

## ClusterTimeSearcher
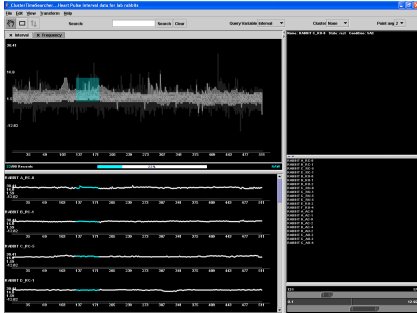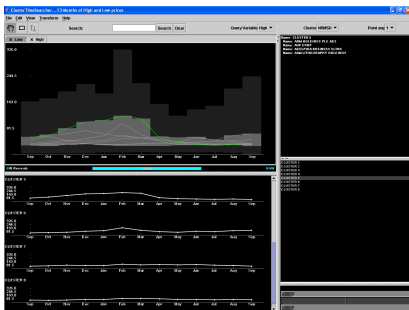


## ClusterTimeSearcher



## ClusterTimeSearcher



## Results

- No user input, so it is largely based on intuition
- Somewhat slower than original TimeSearcher
- Poor precision of the original is not good for complex numerical representations.
- Tradeoff: speed vs. precision vs. being able to represent the value on the screen
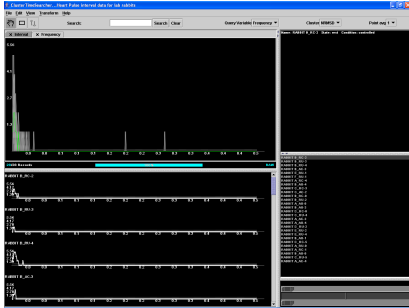
## Results



## Results

- We observe that ClusterTimeSearcher is good for any kind of data, not just heart data
- From experimentation, number of clusters is largely subjective
- To make a useful classification, we need to start with as much info about a condition as we can

## Results



## Results

- Clustering heart data based on power spectrum can provide the initial intuition
- Suspicion: from the experimentation on the data used, it is the short, abrupt changes in the signal that hold the most information
- Cluster count too low = we lose the details we are trying to find