A Guide to Open-Access Databases and Open-Source Software on PhysioNet

George B. Moody
Harvard-MIT Division of Health Sciences and Technology
Cambridge, Massachusetts, USA

Outline

- Background
- Open-Access Databases
- Open-Source Software

What is PhysioNet?

- A web site (http://www.physionet.org)
- A public service of the NIH-sponsored Research Resource for Complex Physiologic Signals (MIT, Harvard, BU, McGill), established in 1999
- A source of freely available physiologic data and open-source software to support the biomedical research community

PhysioBank: Open-Access Databases

- Well over 4000 recordings of digitized physiologic signals and time series
- Organized in over 40 databases (collections of recordings)
- Most recordings are annotated (significant features have been marked to a precision of a sampling interval)

Database categories

- Multi-parameter (ECG with other signals)
- ECG
- Interbeat (RR) intervals
- Others (gait, etc.)

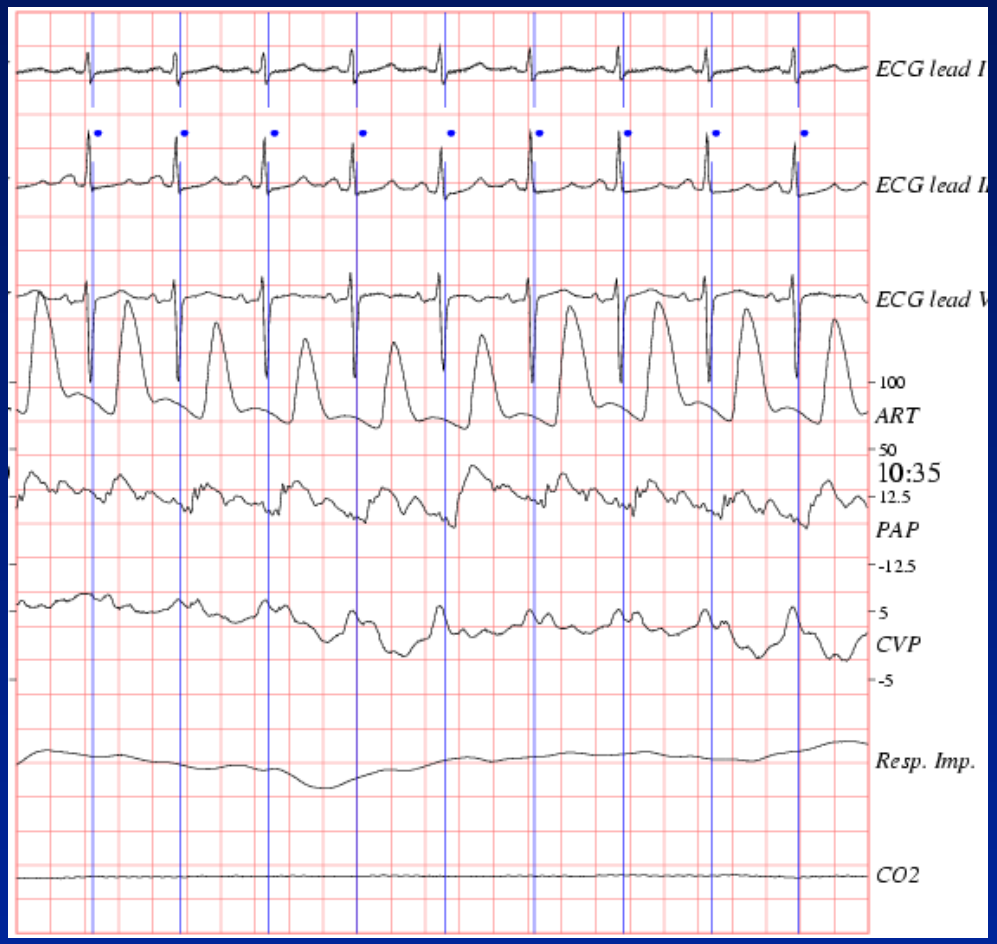
Multi-parameter databases

- MGH/MF Waveform Database: 250 records

3 ECG signals, ABP, PAP, CVP, respiration, CO2, others; from ICU monitors

About 2 hours/record

Each QRS annotated



Multi-parameter databases

- Fantasia Database

  ECG and respiration collected from 40 rigorously-screened healthy volunteers watching *Fantasia*

  Uncalibrated continuous NIBP available for 20 recordings

  Each QRS annotated

Multi-parameter databases

- MIMIC Database

  121 long-duration recordings (many 40 hours or more)

  72 records include up to seven signals (ECG, ABP, etc.) as available from ICU monitors in each case

  All records include monitors' measurements ("numerics": HR, respiration rate, etc.) and alarms

Follow-up MIMIC 2 in progress: records up to several weeks long, ~3000 with signals, ~17000 with numerics

Multi-parameter databases

- Apnea-ECG Database
- MIT-BIH Polysomnographic Database
- Sleep-EDF Database

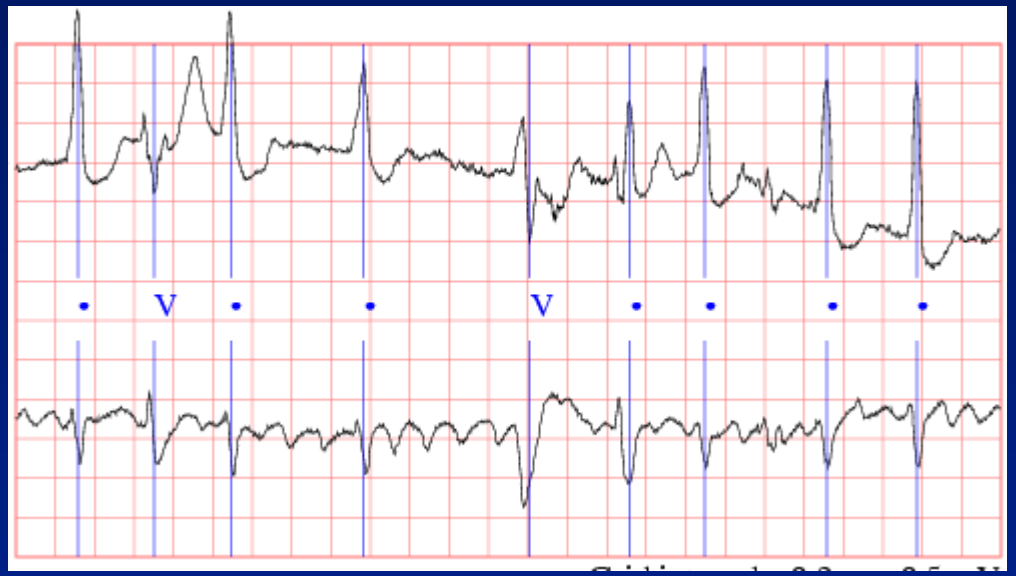
ECG Databases

- MIT-BIH Arrhythmia Database

48 half-hour records

2 ECG signals

Each QRS annotated



Completed in 1980, distributed on tape and CD to over 500 researchers and device developers worldwide

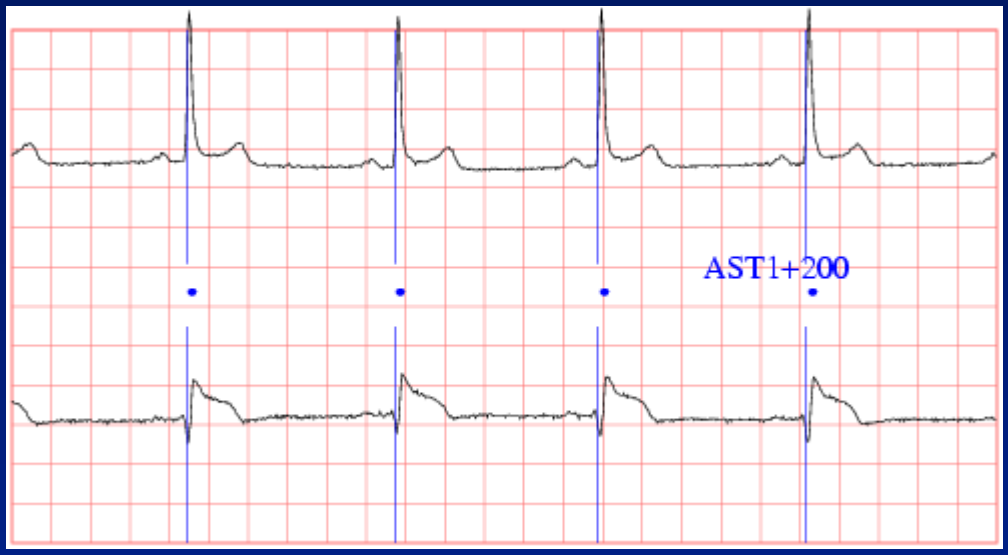
ECG Databases

- European ST-T Database

  90 two-hour records
  (48 freely available)

  Each beat annotated

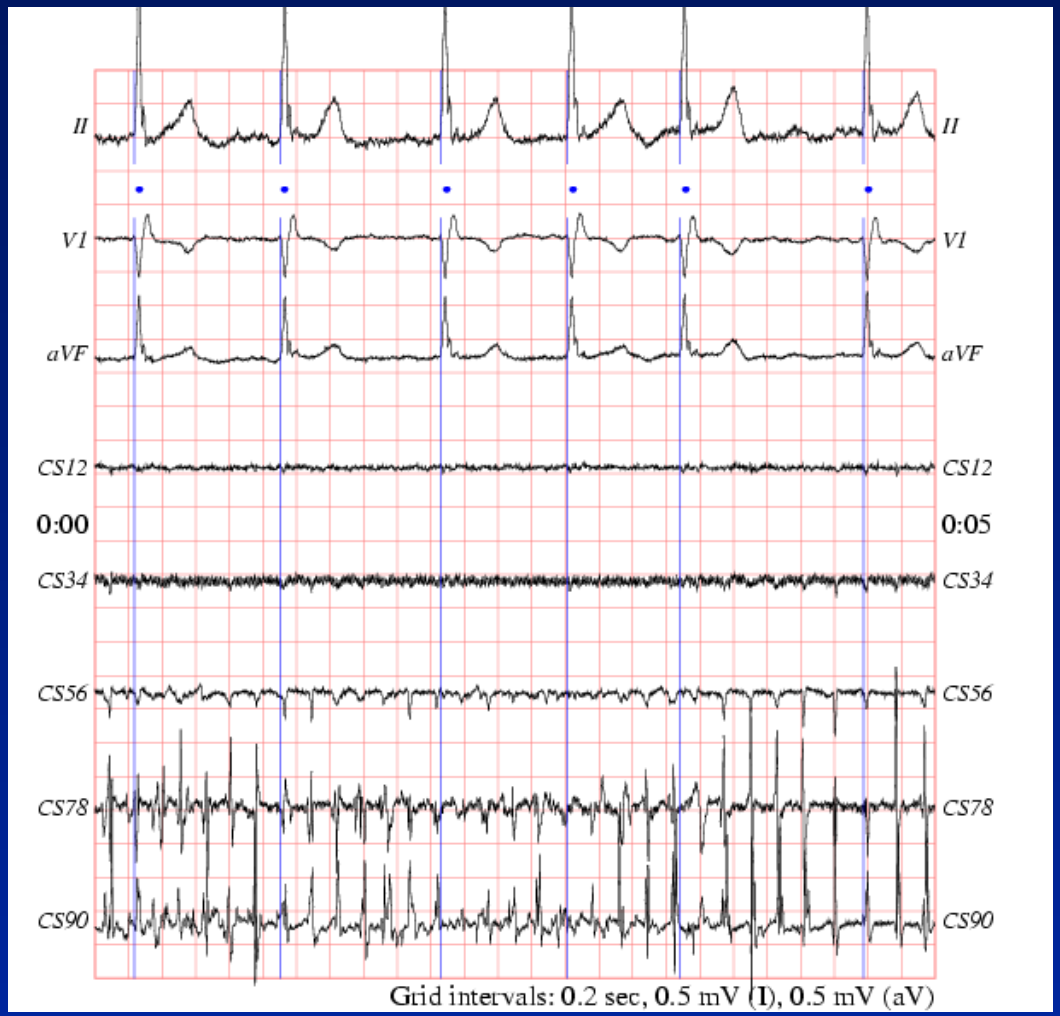
  ST changes annotated



ECG databases

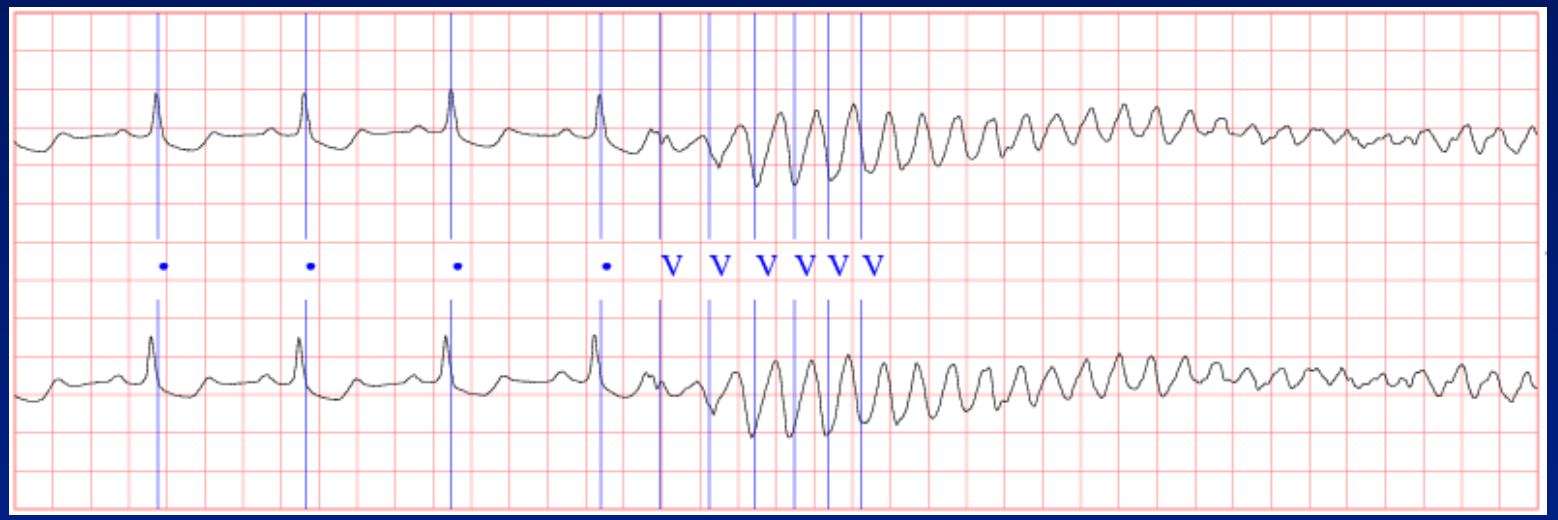
- Intracardiac Atrial Fibrillation Database

8 records, each containing 3 surface ECG signals and 5 intracardiac signals



ECG databases

- Sudden Cardiac Death Holter Database



23 records; contributions of additional records and of annotations (currently complete for 12 records) sought

ECG databases

- PTB Diagnostic ECG Database

549 records from 240 patients and 54 healthy subjects

Most records two minutes long

1000 samples/sec, 16-bit resolution, 15 signals

ECG databases

- Long-Term ST Database
- MIT-BIH Noise Stress Test Database
- BIDMC Congestive Heart Failure Database
- Post-Ictal Heart Rate Oscillations in Partial Epilepsy
- QT Database
- AF Termination Challenge Database

ECG Databases

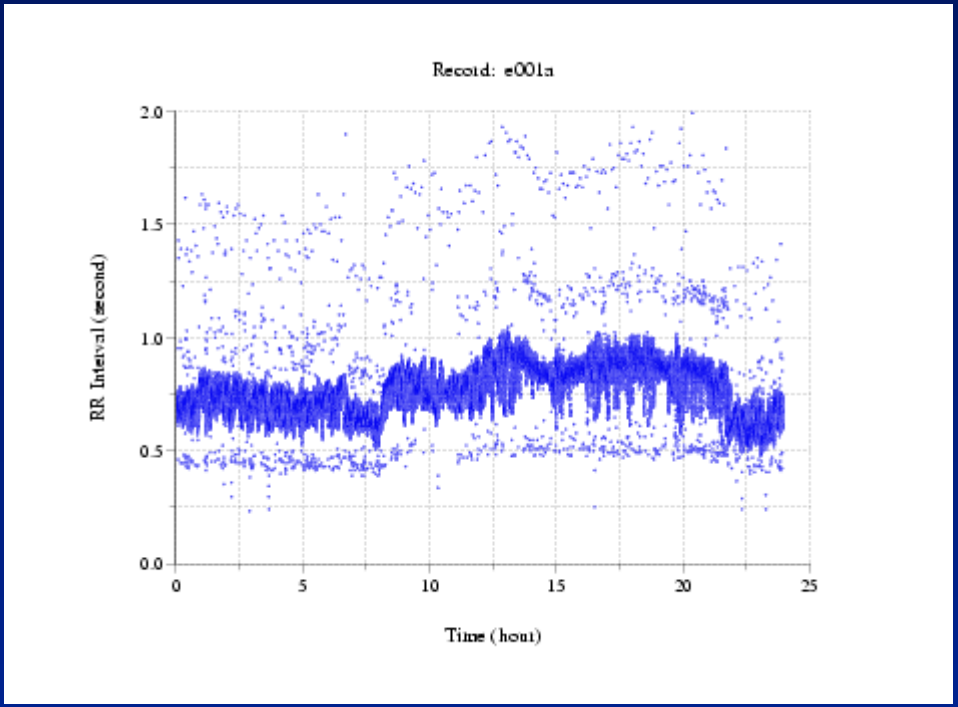
- Creighton University Ventricular Tachyarrhythmia Database
- MIT-BIH Atrial Fibrillation Database
- MIT-BIH ECG Compression Test Database
- MIT-BIH Long-Term Database
- MIT-BIH ST Change Database
- ... and more

Interbeat (RR) interval Databases

- CAST RR Interval Sub-Study Database

1543 beat annotation files (no signals) from 809 subjects entered in the CAST study

Each record is about 24 hours



Total number of beat annotations: about 150 million

Interbeat (RR) interval databases

- Congestive Heart Failure RR Interval Database
- Exaggerated heart rate oscillations during two meditation techniques
- Normal Sinus Rhythm RR Interval Database

Open-source software (PhysioToolkit)

- Data visualization
- Importing and exporting data
- Signal and time series analysis
- Models and simulations
- Software development tools

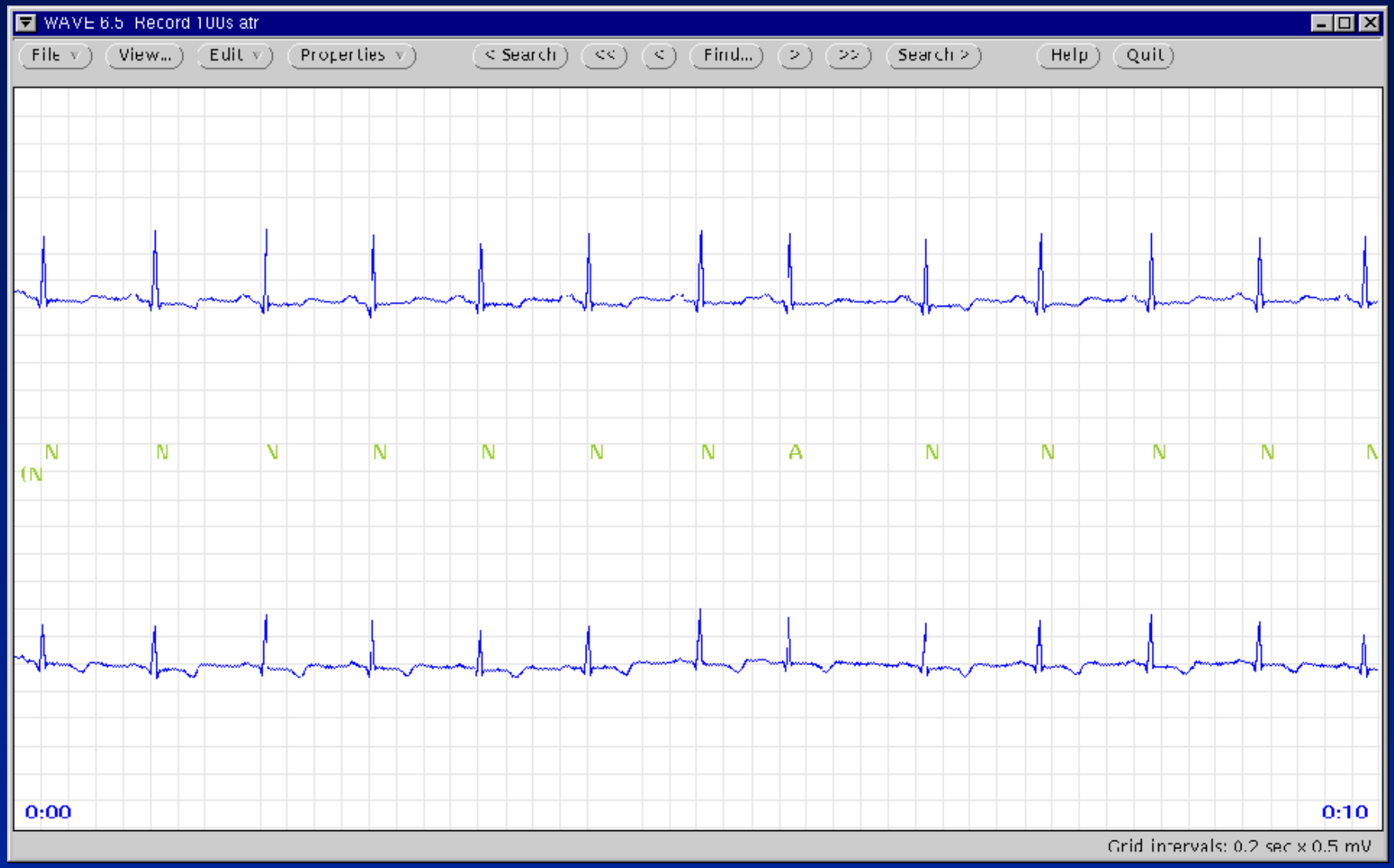
WFDB Software Package

- >80 application programs for processing and analyzing physiologic signals
- Common library for reading and writing signals and annotations
- ANSI/ISO C code portable across all popular platforms (GNU/Linux, Mac OS X, MS-Windows, Unix)
- Free (open-source) software

Examples of WFDB applications

- QRS and pulse detectors
- Heart rate variability analysis
- Digital filters
- Annotation-triggered signal averaging
- ECG-derived respiration
- WAVE (waveform analyzer/viewer/editor)

WAVE


WAVE 6.5 Record 100s atr
File
View...
Edit
Properties
< Search
<<
<
Find...
>
>>
Search >
Help
Quit
N N V N N N N A N N N N N
(N
0:00
0:10
Grid intervals: 0.2 sec x 0.5 mV

WAVE

- Interactive environment for analyzing, viewing, and editing (annotating) signals
- Used to support development of many waveform databases for research
- WAVE processes running on the same or different computers can communicate and synchronize their displays, enabling collaborative annotation and data review across the Internet

WFDB library

- Reads and writes digitized signals and annotations in dozens of formats
- WFDB library provides uniform access to data, finding input data using a path search, and translating among all supported formats
- Applications using the WFDB library don’t need to know how or where data are stored

What is “web-enabled” software?

- Web-enabled software can read and process data from web servers in the same way as data from local files
- Web browsers are familiar examples of web-enabled software, but ...
- Web browsers are not the only means of retrieving data from web servers!

Why?

Why would anyone want web-enabled physiologic signal processing software?

- Clinicians can seek and provide expert interpretation of physiologic signals in telemedicine applications
- Researchers can develop and use shared archives of signals, such as those available via PhysioNet

Using the WFDB library: an example

- Open two signals from a record named `mitdb/200`:

```
isigopen("mitdb/200", s, 2);
```

- Skip the first 10 minutes and 31 seconds:

```
isigsettime("10:31");
```

- Read and print 5 samples of each signal:

```
for (i = 0; i < 5; i++) {
     getvec(v);
     printf("%d %d\n", v[0], v[1]);
}
```

Using the WFDB library: an example

```
#include <wfdb/wfdb.h>

main() {
      int i, v[2];
      WFDB_Siginfo s[2];

      isigopen("mitdb/200", s, 2);
      isigsettime("10:31");
      for (i = 0; i < 5; i++) {
            getvec(v);
            printf("%d %d\n", v[0], v[1]);
      }
}
```

Using the WFDB library: an example

- Compile the example:

```
cc -o example example.c -lwfdb
```

- Run the example and collect output:

```
./example
```

```
942 965
942 963
940 958
939 958
936 957
```

What does the example demonstrate?

- Only the record name is needed in order to read the signals
- Samples are readable without needing to know how they are stored
- A trivial task can be accomplished using a trivial program
- The WFDB library hides the complexity of finding inputs and interpreting data formats

Open-source software on PhysioNet

- 600 to 900 users download the WFDB software package from PhysioNet each month
- 15% to 20% of all data downloaded from PhysioNet are retrieved by web-enabled WFDB applications
- An active user community contributes new applications and improvements to existing applications

Visit PhysioNet!

http://physionet.org/