



# A modification of the Smith-Waterman algorithm to align sequences of numeric data

Marcus Vollmer

Institute of Bioinformatics, University Medicine Greifswald, Germany  
German Centre for Cardiovascular Research (DZHK), partner site Greifswald

**25 July 2022, Statistical Computing, Günzburg**

1 Motivation

2 Smith-Waterman modification

# 1. Motivation

For heart rate analysis an ECG is non-invasively measuring heart activity.

## Research questions:

- ▶ Is expensive medical equipment superior to consumer products in terms of signal quality?
- ▶ How does the quality change with physical and mental workload?
- ▶ Which devices are superior in terms of usability?

For heart rate analysis an ECG is non-invasively measuring heart activity.

## Research questions:

- ▶ **Is expensive medical equipment superior to consumer products in terms of signal quality?**
- ▶ How does the quality change with physical and mental workload?
- ▶ Which devices are superior in terms of usability?

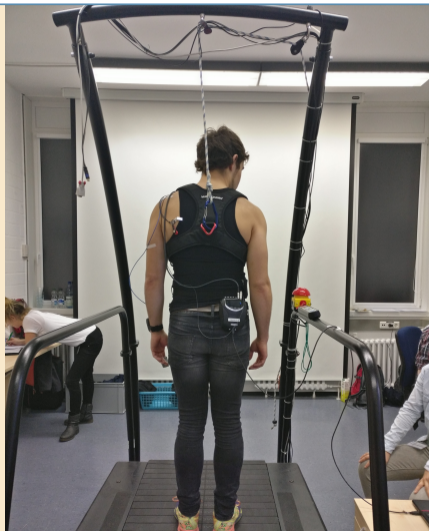
For heart rate analysis an ECG is non-invasively measuring heart activity.

## Research questions:

- ▶ **Is expensive medical equipment superior to consumer products in terms of signal quality?**
- ▶ How does the quality change with physical and mental workload?
- ▶ Which devices are superior in terms of usability?

Signal quality can be expressed by the ability of standard algorithms to detect the R peaks from the ECGs.

# Experimental Setup



- ▶ 5 min standing rest
- ▶ 5 min walking on treadmill (1.2 m/s)
- ▶ cognitive test (2-back audio test)
- ▶ 5 min walking on treadmill (1.2 m/s, 15% gradient)

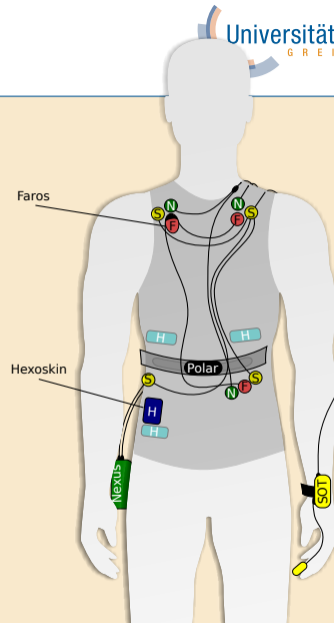
In between: NASA Task Load Index to measure individual strain

Conference paper 2019 [3]


Data set 2020 [4, 1] [▶ PhysioNet](#)

# Experimental Setup

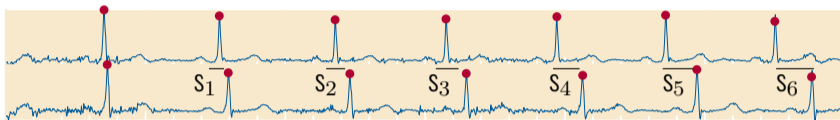
- ▶ eMotion Faros 360° | 1000 Hz
- ▶ SOMNOtouch NIBP | 512 Hz
- ▶ NeXus-10 MKII | 8000 Hz
- ▶ Polar RS800 Multi | 1000 Hz
- ▶ Hexoskin Smart Shirt Hx1 | 256 Hz





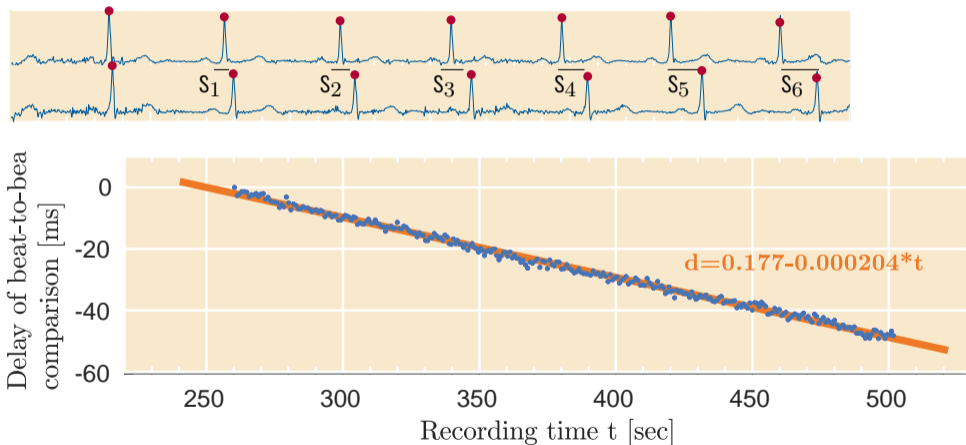
	eMotion Faros 360°	Hexoskin Hx1	NeXus-10 MKII	Polar Multi RS800	SOMNOtouch NIBP
					
Follow-up	2017 Bittium Faros	2015 Hexoskin Smart Shirt	2016 -	2007 Polar Watch V800 + Polar H10	2015 -
Supplies	ECG electrodes	Skin preparation gel	ECG electrodes	-	ECG electrodes
Sampling rate	1000 Hz	256 Hz	8000 Hz	1000 Hz	512 Hz

# Linear Adjustment of Sampling Frequencies

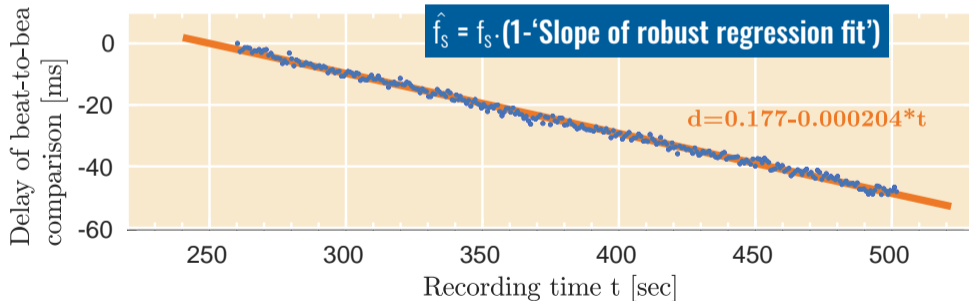
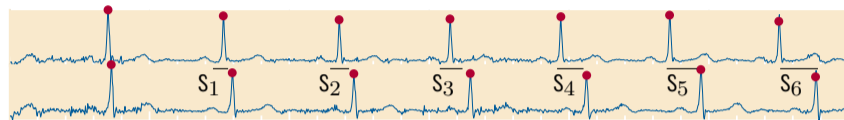


- ▶ Pairwise differences  $s_i$  are linear decreasing or increasing with time

# Linear Adjustment of Sampling Frequencies



# Linear Adjustment of Sampling Frequencies



# Linear Adjustment Factors

Manufacturer's specifications and actual sampling frequency (Hexoskin served as a reference with constant  $f_s$ ).  
Two Faros' devices in use,  $\hat{f}_s$  splitted according to device ID.

Recording Device	$f_s$	Mean $\hat{f}_s$	min	max
SomnoTOUCH NIBP	512	511.97	511.97	511.97
NeXus-10 MKII	8000	7999.67	7999.67	7999.68
eMotion Faros 360°, ID1	1000	1000.29	1000.19	1000.36
eMotion Faros 360°, ID2	1000	1000.18	1000.15	1000.21
Polar RS800 Multi	1000	999.91	999.87	999.95

# Linear Adjustment Factors

Manufacturer's specifications and actual sampling frequency (Hexoskin served as a reference with constant  $f_s$ ).  
Two Faros' devices in use,  $\hat{f}_s$  splitted according to device ID.

Recording Device	$f_s$	Mean $\hat{f}_s$	min	max
SomnoTOUCH NIBP	512	511.97	511.97	511.97
NeXus-10 MKII	8000	7999.67	7999.67	7999.68
eMotion Faros 360°, ID1	1000	1000.29	1000.19	1000.36
eMotion Faros 360°, ID2	1000	1000.18	1000.15	1000.21
Polar RS800 Multi	1000	999.91	999.87	999.95

constant  
constant

# Linear Adjustment Factors

Manufacturer's specifications and actual sampling frequency (Hexoskin served as a reference with constant  $f_s$ ).  
Two Faros' devices in use,  $\hat{f}_s$  splitted according to device ID.

Recording Device	$f_s$	Mean $\hat{f}_s$	min	max
SomnoTOUCH NIBP	512	511.97	511.97	511.97
NeXus-10 MKII	8000	7999.67	7999.67	7999.68
eMotion Faros 360°, ID1	1000	1000.29	1000.19	1000.36
eMotion Faros 360°, ID2	1000	1000.18	1000.15	1000.21
Polar RS800 Multi	1000	999.91	999.87	999.95

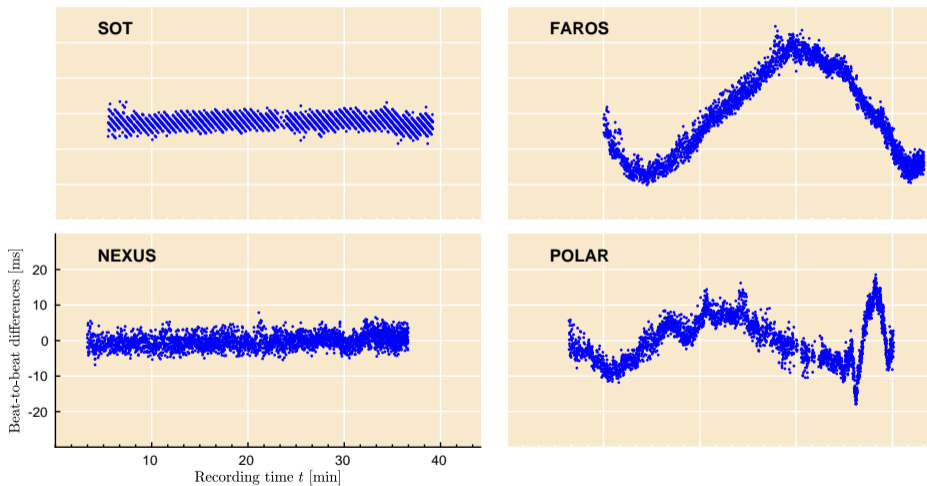
constant  
constant

fluctuations

fluctuations

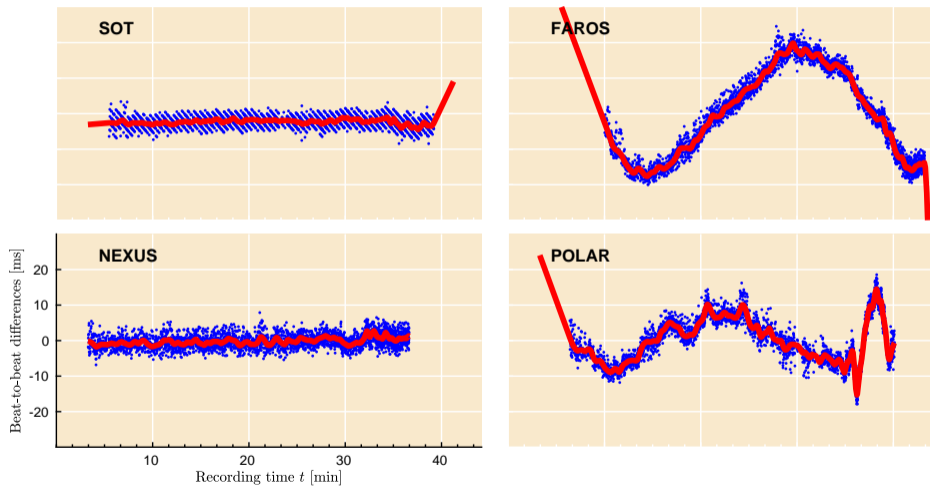
fluctuations

# Is a linear adjustment sufficient?





# Is a linear adjustment sufficient?



## 2. Smith-Waterman modification

## Signal quality based on beat detectability

RR intervals is the sequence of the durations between successive R peaks of the heart cycle. Any error in these intervals is the reflection of errors in the beat detection.

### **Alignment and comparison of RR intervals:**

- ▶ Polar RR interval sequences were generated by an in-house algorithm
- ▶ RR intervals were generated for all other devices using a standard detector applied to the ECG
- ▶ Reference intervals (ground truth) were generated by experienced researchers based on all signals

## Signal quality based on beat detectability

RR intervals is the sequence of the durations between successive R peaks of the heart cycle. Any error in these intervals is the reflection of errors in the beat detection.

### Alignment and comparison of RR intervals:

- ▶ Polar RR interval sequences were generated by an in-house algorithm
- ▶ RR intervals were generated for all other devices using a standard detector applied to the ECG
- ▶ Reference intervals (ground truth) were generated by experienced researchers based on all signals

**Is there an efficient method to align and evaluate the quality of RR intervals?**

## Relation to sequence alignment

Smith-Waterman [2] published a well-known algorithm to perform **local sequence alignment of nucleic acid sequences or protein sequences**.

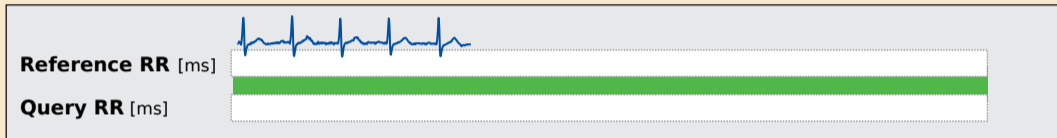
```
Sequence 1: AAG-GCCAAGGAG
             ..|  |||  .||
Sequence 2: TTGTGCC-TGG
```

The alignment is composed of **matches**, **mismatches** (mutations) and **gaps** (insertions, deletions). The overall score of an alignment is a sum of individual scores (match=10, mismatch=-2, gap=-7):

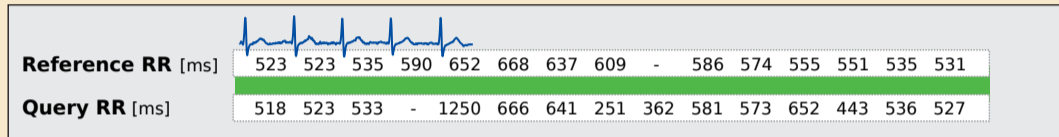
$$\text{Score} = 6 * 10 + 3 * -2 + 2 * -7 = 40$$

Dynamic programming is used in the matching process to find the optimal local alignment with respect to the scoring system being used. [▶ Wiki](#)

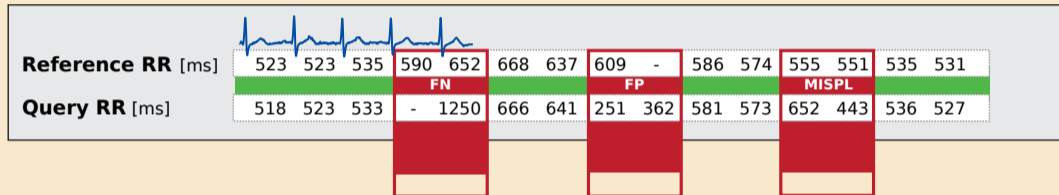
# Smith-Waterman modification



# Smith-Waterman modification

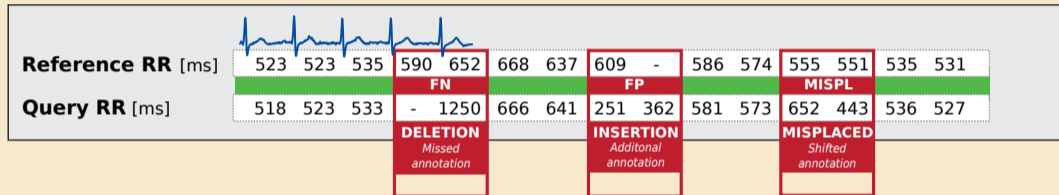


# Smith-Waterman modification

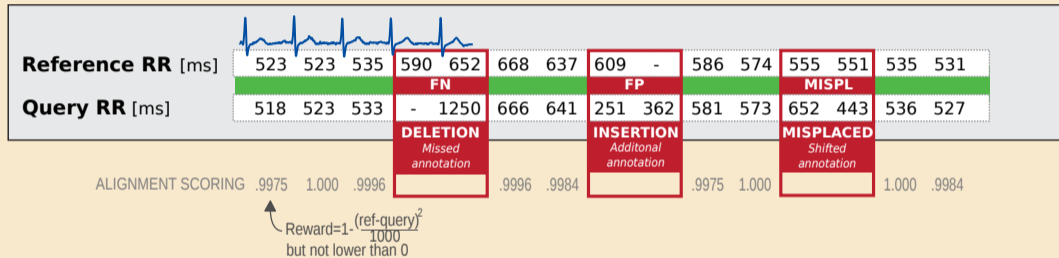




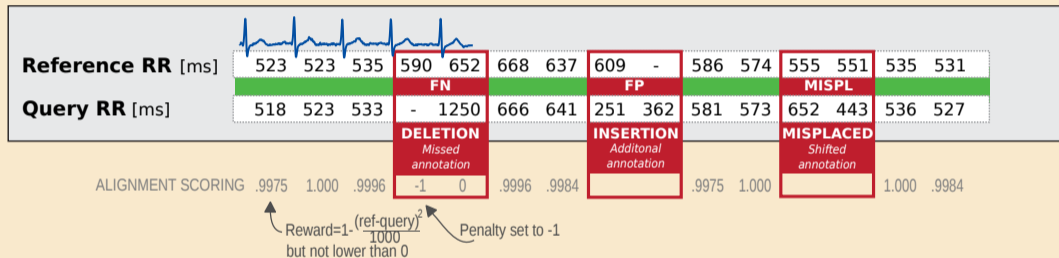
# Smith-Waterman modification



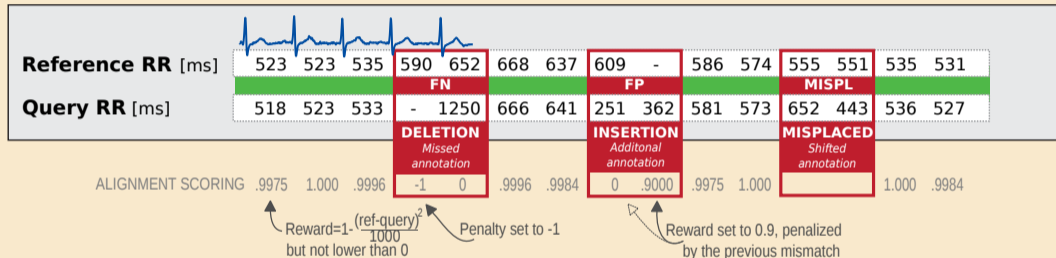
## Smith-Waterman modification



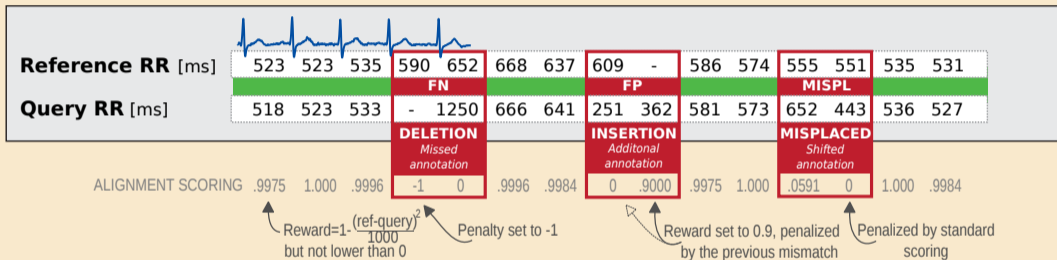
# Smith-Waterman modification



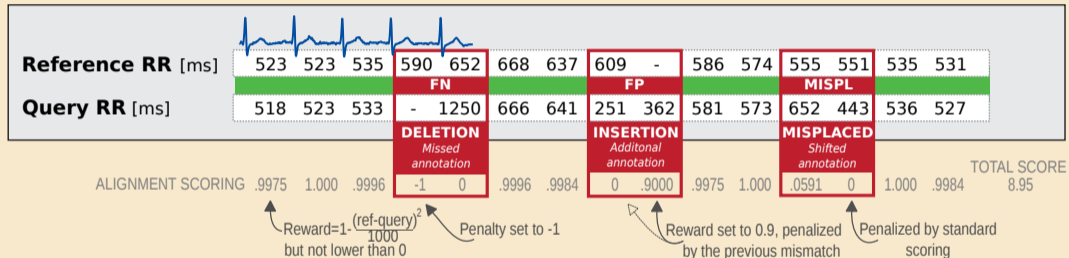
# Smith-Waterman modification



# Smith-Waterman modification



# Smith-Waterman modification



## Scoring definition for RR interval alignment

Eq 1 and Eq 2 shows the definition of the scoring matrix  $H$  of sequences  $a$  and  $b$  having  $m$  and  $n$  values respectively. Function  $s$  of Eq 3 defines the reward of matched RR intervals allowing for minor deviations (reward=1 for 0 ms difference, reward=0.9 for  $\pm 10$  ms difference, reward=0.6 for  $\pm 20$  ms difference):

$$\begin{aligned}
 H(i, 0) &= 0, & i &= 0, \dots, m \\
 H(0, j) &= 0, & j &= 0, \dots, n
 \end{aligned} \tag{1}$$

$$H(i, j) = \max \left\{ \begin{array}{ll} 0 & \text{No similarity} \\ H(i-1, j-1) + s(a_i, b_j) & \text{Match/Mismatch} \\ H(i-1, j) - 1.0 & \text{Deletion} \\ H(i, j-1) + 0.9 & \text{Insertion} \end{array} \right\}, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{2}$$

$$s(a_i, b_j) = \max \left\{ \begin{array}{ll} 0 & \text{capped score} \\ 1 - \frac{(a_i - b_j)^2}{1000} & \text{RR interval deviation} \end{array} \right\}, \quad a_i \text{ and } b_j \text{ in ms} \tag{3}$$

Adapted code from Needleman-Wunsch implementation:

- ▶ Source code: <https://github.com/meicaljohnjones/Needleman-Wunsch>
- ▶ Author: Michael John Jones



# Beat detection performance

	Overall	P1 – Rest	P2 – Walking	P3 – 2-back	P4 – Uphill
total reference beats	42364	5623	6235	6106	9222
avg. peak heart rate	–	93.6 bpm	103.3 bpm	105.2 bpm	153.7 bpm
avg. median breathing rate	–	14.5 min <sup>-1</sup>	24.1 min <sup>-1</sup>	24.3 min <sup>-1</sup>	27.1 min <sup>-1</sup>
<b>FP FN Misplaced using Smith-Waterman approach</b>					
eMotion Faros 360°	5 0 19	0 0 0	0 0 0	0 0 0	0 0 6
Hexoskin Hx1	85 10 765	3 0 5	114 0 204	11 0 14	29 1 270
NeXus-10 MKII	40 5 765	0 0 0	3 0 4	15 0 40	0 1 309
SOMNOtouch NIBP	63 0 202	0 0 0	8 5 134	0 0 2	0 0 65
Polar RS800 Multi	28 29 333	0 3 18	6 0 57	3 3 66	0 0 11

**Table:** QRS detection performance through RR interval alignment using the modified Smith-Waterman algorithm for numericals. Errors are expressed by false positive counts (FP), false negative counts (FN), and the number of misplaced annotations from unsynchronized data (50 ms tolerance).

## **We are able to rate the performance and to classify beat detection error:**

- ▶ Insertions represent RR intervals that are splitted into smaller sub-intervals due to a false positive beat location
- ▶ Deletions can be interpreted as missing beat locations
- ▶ Mutations represent inaccurate RR intervals due to misplaced beat locations

The SW-approach represents an efficient way of quantifying errors due to poor signal quality and weaknesses in heartbeat detection.

## **Efficiency is expressed by:**

- ▶ No need to synchronize signals or frequency correction
- ▶ Automated handling of alterations of sampling frequencies during long-term measurements
- ▶ Signals are allowed to interrupt or stop for a period



**Thank you for your  
time and attention**

**I am happy to answer questions**



A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley.

PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.

*Circulation*, 101(23):e215–e220, 2000.



T. F. Smith, M. S. Waterman, et al.

Identification of common molecular subsequences.

*Journal of Molecular Biology*, 147(1):195–197, 1981.



M. Vollmer, D. Bläsing, and L. Kaderali.

Alignment of multi-sensored data: adjustment of sampling frequencies and time shifts.

In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.



M. Vollmer, D. Bläsing, J. E. Reiser, M. Nisser, and A. Buder.

Simultaneous physiological measurements with five devices at different cognitive and physical loads (1.0.0) [data set].

PhysioNet, <https://doi.org/10.13026/CHD5-T946>, 2020.